SCIENCE LIBRARY

DEC 2 - 1932

Vol. XV, No. 6 *PSYCHOLOGICAL REVIEW PUBLICATIONS* DECEMBER, 1932

Journal of Experimental Psychology

EDITED BY

SAMUEL W. FERNBERGER, UNIVERSITY OF PENNSYLVANIA

HOWARD C. WARREN, PRINCETON UNIVERSITY (*Review*)
WALTER S. HUNTER, CLARK UNIVERSITY (*Index*)
HERBERT S. LANGFELD, PRINCETON UNIVERSITY (*Monographs*)
EDWARD S. ROBINSON, YALE UNIVERSITY (*Bulletin*)

CONTENTS

A Disproof of the Law of Effect and a Substitution of the Laws of Emphasis, Motivation and Disruption: E. C. TOLMAN, C. S. HALL AND E. P. BRETNALL 601

The Prediction of Vocational Aptitude and Success from Photographs: M. S. VITELES AND K. R. SMITH 615

The Functional Significance of Imagery Differences: F. C. DAVIS 630

The Comparative Retention Values of a Maze Habit, of Nonsense Syllables, and of Rational Learning: J. A. MCGEOCH 662

Mental Set in Relation to Retroactive Inhibition: O. P. LESTER 681

Habituation and Perseverational Characteristics of Two Forms of Indirect Suggestion: C. L. HULL AND M. C. FORSTER 700

Factors Determining the Relative Efficacy of the Whole and Part Methods of Learning: A. J. DAVIS AND M. MEENES 716

Memory for Pleasant, Unpleasant, and Indifferent Pairs of Words: H. CASON AND F. C. LUNGREN 728

Equal Weights and Psychophysical Judgments: L. ARONS AND F. W. IRWIN 733

Influence of an Interpolated Electric Shock upon Recall: M. M. WHITE 752

Ocular Dominance in Young Children: R. UPDEGRAFF 758

The Autonomic Nervous System as a Factor in the Psychogalvanic Reflex: W. D. O'LEARY 767

The Regulated Alternating Current as a Time Measure: T. H. HOWELLS 773

A Flexible Apparatus for Recording Reading Reactions: M. A. TINKER 777

PUBLISHED BI-MONTHLY
FOR THE AMERICAN PSYCHOLOGICAL ASSOCIATION
BY THE PSYCHOLOGICAL REVIEW COMPANY
PRINCE AND LEMON STS., LANCASTER, PA.
AND PRINCETON, N. J.

Entered as second-class matter, July 1, 1920, at the post-office at Lancaster, Pa.

PUBLICATIONS

OF THE

AMERICAN PSYCHOLOGICAL ASSOCIATION

EDITED BY

HOWARD C. WARREN, PRINCETON UNIVERSITY (*Review*)
S. W. FERNBERGER, UNIVERSITY OF PENNSYLVANIA (*J. Exper. Psych.*)
WALTER S. HUNTER, CLARK UNIVERSITY (*Index and Abstracts*)
HENRY T. MOORE, SKIDMORE COLLEGE (*J. Abn. and Soc. Psychol.*)
HERBERT S. LANGFELD, PRINCETON UNIVERSITY (*Monographs*)
EDWARD S. ROBINSON, YALE UNIVERSITY (*Bulletin*)

HERBERT S. LANGFELD, Business Editor

PSYCHOLOGICAL REVIEW

containing original contributions only, appears bi-monthly, January, March, May, July, September, and November, the six numbers comprising a volume of about 540 pages.

PSYCHOLOGICAL BULLETIN

containing critical reviews of books and articles, psychological news and notes, university notices, and announcements, appears monthly (10 numbers), the annual volume comprising about 720 pages. Special issues of the BULLETIN consist of general reviews of recent work in some department of psychology.

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

containing original contributions of an experimental character, appears bi-monthly, February, April, June, August, October, and December, the six numbers comprising a volume of about 700 pages (from Jan. 1, 1932).

PSYCHOLOGICAL INDEX

is a compendious bibliography of books, monographs, and articles upon psychological and cognate topics that have appeared during the year. The INDEX is issued annually in June, and may be subscribed for in connection with the periodicals above, or purchased separately.

PSYCHOLOGICAL ABSTRACTS

appears monthly, the twelve numbers and an index supplement making a volume of about 700 pages. The journal is devoted to the publication of non-critical abstracts of the world's literature in psychology and closely related subjects.

PSYCHOLOGICAL MONOGRAPHS

consists of longer researches or treatises or collections of laboratory studies which it is important to publish promptly and as units. The price of single numbers varies according to their size. The MONOGRAPHS appear at irregular intervals and are gathered into volumes of about 500 pages.

JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY

appears quarterly, April, July, October, January, the four numbers comprising a volume of 448 pages. The journal contains original contributions in the field of abnormal and social psychology, reviews, notes and news.

ANNUAL SUBSCRIPTION RATES

Review: $5.50 (Foreign, $5.75). Index: $4.00 per volume.
Journal: $7.00 (Foreign, $7.25). Monographs: $6.00 per volume (Foreign, $6.30).
Bulletin: $6.00 (Foreign, $6.25). Abstracts: $6.00 (Foreign, $6.25).
Abnormal and Social: $5.00 (Foreign, $5.25). Single copies $1.50.
Current numbers: Journal, $1.25; Review, $1.00; Abstracts, 75c; Bulletin, 60c.

COMBINATION RATES (from Jan. 1, 1932)

Review and Bulletin: $10.00 (Foreign, $10.50).
Review and J. Exp.: $11.00 (Foreign, $11.50).
Bulletin and J. Exp.: $12.00 (Foreign, $12.50).
Review, Bulletin, and J. Exp.: $16.00 (Foreign, $16.75).
Review, Bulletin, J. Exp., and Index: $19.00 (Foreign, $19.75).

Subscriptions, orders, and business communications should be sent to the

PSYCHOLOGICAL REVIEW COMPANY

PRINCETON, NEW JERSEY

Journal of
Experimental Psychology

VOL. XV, No. 6 DECEMBER, 1932

A DISPROOF OF THE LAW OF EFFECT AND A SUBSTITUTION OF THE LAWS OF EMPHASIS, MOTIVATION AND DISRUPTION

BY E. C. TOLMAN, C. S. HALL, AND E. P. BRETNALL

University of California

The present investigation was formulated with the idea of discovering whether or not the Law of Effect in the form in which it is usually stated by Thorndike (for example (**2**), p. 36) necessarily and inevitably holds.[1] The Law states that, in a trial and error situation, the right responses tend to get selected as against the wrong responses primarily because the right responses are followed by a good effect (*i.e.*, satisfiers) whereas the wrong responses tend to drop out because they are followed by bad effects (*i.e.*, annoyers). It was the aim of the present experiment to discover what would happen if an annoyer (electric shock) were applied not, as is usually done, after the wrong responses but after the right responses. We sought, that is, to compare the results for such an arrangement with those for the more usual set-up in which the annoyer (electric shock) was applied after the wrong responses.

Apparatus.—The apparatus used was a device intended to be for human beings relatively comparable to a maze for rats. We have called it the 'punchboard maze.'[2] It consisted of

[1] A brief preliminary statement of the experiment has already appeared. See Tolman (**3**), pp. 344–345.

[2] Designed by Robert H. Bruce and hitherto unreported.

thirty pairs of holes arranged in an irregular spatial order on a board three feet square (see Fig. 1). The subject's task was to go through the thirty pairs, punching one hole in each pair with a stylus (analogous to entering the blind or the true path section at each of thirty choice points in a maze) and thus to

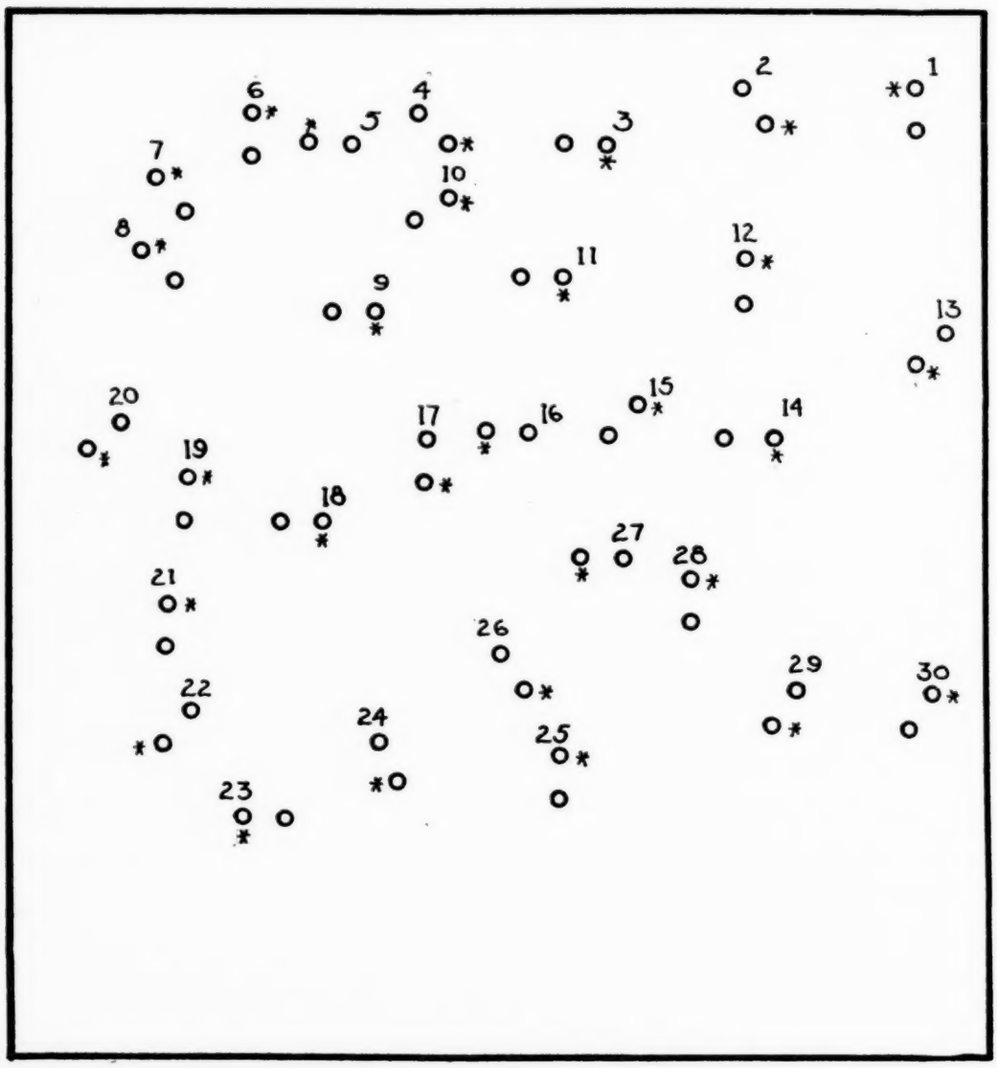


FIG. 1. Punch Board Maze. The starred holes are those which for Groups I, II, III and IV sounded. The stars did not appear on the actual maze itself.

learn for each pair which of the two holes was correct. The board was placed on an easel so that it sloped at a slight angle away from the subject and so that the bottom edge was twenty-one inches from the floor. Some of the subjects stood and some sat during the test performance. The board had 60 holes and was covered with a facing of heavy wrapping

paper. The holes were three-quarters of an inch deep and were backed by uniform brass contact plates. The contact plates of one hole in each of the thirty pairs was connected with an electric circuit so that when the stylus was introduced it rang a buzzer or, in the case of some groups, gave an electric shock plus the ringing of the buzzer.

Subjects and Procedure.—There were five groups of subjects. The conditions under which these different groups were tested were as follows:

Groups I and II and I*a* were the control groups. For them no shock but merely the electric buzzer was used. This buzzer was connected with one hole of each of the pairs of holes so that when the metal stylus was pushed into the hole against the contact plate the buzzer rang.

Group I learned to *avoid* these holes which thus sounded the buzzer when the stylus was pushed into them. This group will be referred to as the Bell-Wrong group (BW).

Group II learned to *select* the buzzing holes. This group will be referred to as the Bell-Right group (BR).

Group Ia was a special control group. For it the wiring was reversed so that the buzzer sounded on the opposite hole of each pair from what it did for all the other groups. The subjects of Group I*a* were like those of Group I in that they learned to *avoid* the buzzing holes (BW*a*) but the holes which group I*a* avoided were the alternates of those avoided by Group I.

Groups III and IV were the experimental groups. For them the electric shock was added to the buzzer. That is, in the set-up as used for these groups a second circuit was introduced which ran between the bare metal stylus as held by the subject in the one hand and an electrode held by him in his other, passive, hand. This second circuit was the secondary from an inductorium; and the wiring was so arranged that when the stylus made contact with the plate at the back of the hole it caused not only the buzzer to ring as for Groups I and II and I*a* but in addition caused an electric shock to pass through the subject by way of the stylus in his one hand and the electrode in his other.

Group III was like Group I save for the addition of the electric shock. That is, the subjects in this group were required to learn to *avoid* the hole in each pair which both buzzed and gave an electric shock. This group will be referred to as the Bell-Shock-Wrong group (BSW).

Group IV was like Group II save for the addition of the electric shock. That is, the subjects in this group were required to learn to *select* the hole in each pair which both buzzed and gave an electric shock. This group will be referred to as the Bell-Shock-Right group (BSR).

All the subjects were students in the beginning class in psychology. The sizes of the groups and their distributions as to sex are shown in table 1.

TABLE 1

DISTRIBUTION OF SUBJECTS

	Men	Women	Both
I. Bell-Wrong	29	17	46
II. Bell-Right	31	16	47
III. Bell-Shock-Wrong	17	17	34
IV. Bell-Shock-Right	17	17	34
Ia. Bell-Wrong	12	13	25
Totals	106	80	186

The instructions given to all the subjects were as follows: "This is a learning experiment. There are thirty pairs of holes on this board numbered in order from one to thirty. One of each of the pairs of holes rings a buzzer (in the case of Groups III and IV—rings and shocks) and the other hole does not. I want you to place this stylus in each pair of holes from one to thirty. You will have three trials in which you are to hit each hole of each pair. After you have done this, I will tell you whether to learn to hit the ringing or non-ringing series (the ringing-shocking or non-ringing, non-shocking series). You may begin."

The three trials, during which the subject hit all the holes in order, constituted the fore-practice.

After the subject had finished these first three trials, he was then told which series he should learn. He was further

told to correct himself when he made an error by placing the stylus in the right hole of the pair.

When shock was applied, the subject was asked to regulate the inductorium to give the maximum amount of electric current which he felt he could stand. An attempt to offset the factor of adaptation was made by increasing the strength after each trial.

One correct trial was used as the criterion of learning. In a preliminary experiment it was found that a criterion of two correct trials did not affect the results appreciably. If the criterion had not been satisfied by the end of the twentieth trial, the experiment was discontinued. The experiment required about a half-hour of continuous performance.

TABLE 2

MEAN ERRORS

Trial	Group Ia BW*a*	Group I BW	Group II BR	Group III BSW	Group IV BSR
1	11.96	11.91	11.72	13.41	11.18
2	11.96	12.19	11.09	12.26	10.23
3	11.12	11.04	10.49	12.18	10.26
4	10.56	10.85	8.91	11.70	9.73
5	9.64	10.41	8.45	10.20	9.18
6	7.96	8.41	6.81	8.91	7.73
7	8.20	7.96	6.34	8.12	6.91
8	6.56	6.87	5.57	7.91	6.15
9	5.56	5.93	5.40	7.18	5.03
10	4.72	6.00	4.91	5.79	4.35
11	3.48	4.61	4.09	5.59	4.09
12	3.24	4.00	3.57	4.97	3.97
13	3.24	3.28	3.02	4.59	3.12
14	2.56	2.83	2.79	3.59	2.79
15	2.08	2.52	2.21	2.82	2.32
16	1.56	2.09	2.28	2.47	2.06
17	1.00	2.15	2.04	2.41	2.15
18	1.32	1.61	1.68	1.79	1.56
19	1.08	1.39	1.45	1.59	1.50
20	.84	1.13	1.04	1.32	1.15

Results.—Table 2 presents the mean numbers of errors in each group for each of the 20 trials. And figure 2 presents this same data in the form of cumulative error curves [3] for each of the four main groups. Considering these latter it appears at once that the Bell-Right and the Bell-Shock-Right

[3] For the use of cumulative curves as a desirable method of 'smoothing' we are indebted to L. L. Thurstone.

curves are better (*not worse*) than the Bell-Wrong and the Bell-Shock-Wrong curves. It also appears that the difference between Bell-Shock-Right curve and the Bell-Shock-Wrong curve is greater than that between the simple Bell-Right curve and the simple Bell-Wrong curve. And, be it noted, all these

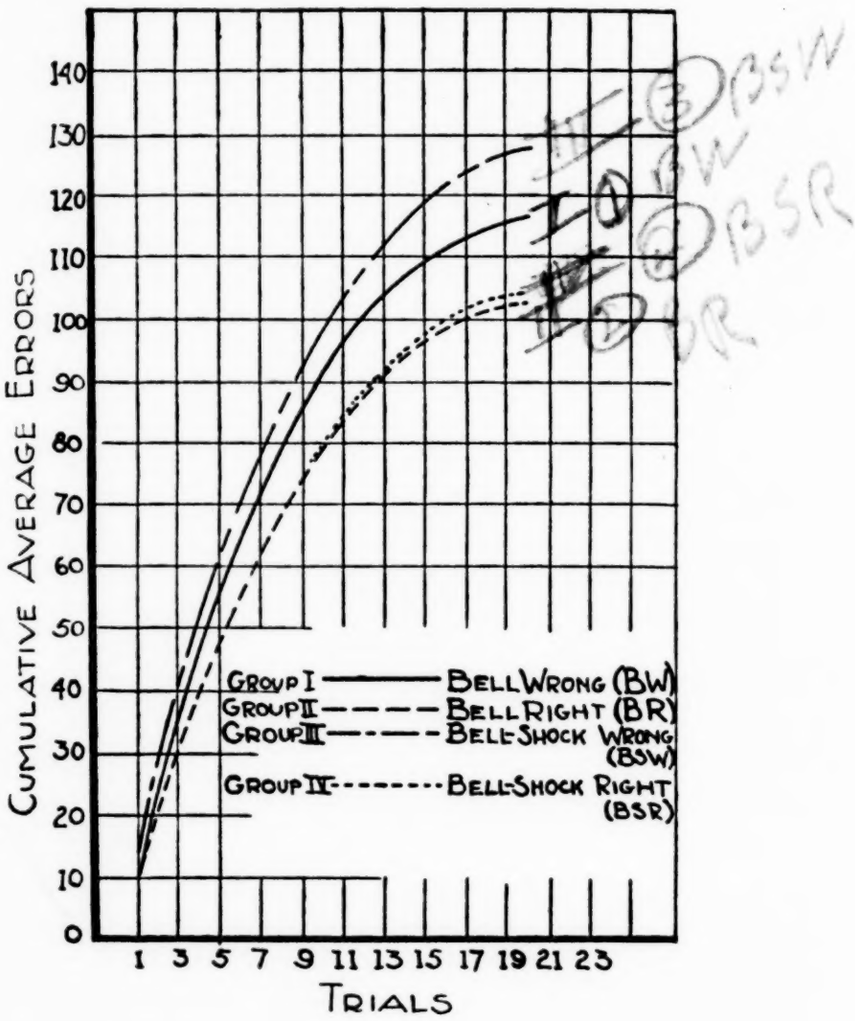


FIG. 2.

facts are contrary to expectation according to any ordinary formulation of the Law of Effect. In fact the only one of the results which suggests any comfort to this usual law is the fact that the Bell-Shock-Right curve is perhaps a little bit worse than the simple Bell-Right curve. In order to be certain of the significances of all these differences we must, however, consider the statistical values, as given in Tables 3 and 4.

From these tables it seems clear that all the differences are

TABLE 3

MEANS, STANDARD DEVIATIONS, AND PROBABLE ERRORS OF THE MEANS FOR THE FOUR MAIN GROUPS

Trials	Bell-Wrong (I)			Bell-Right (II)		
	Mean	S.D.	P.E.$_m$	Mean	S.D.	P.E.$_m$
1– 5.....	56.41	12.00	1.19	50.62	14.52	1.43
6–10.....	35.17	17.87	1.78	29.30	16.45	1.62
11–15.....	17.19	14.63	1.45	15.66	14.19	1.39
16–20.....	8.37	10.82	1.08	8.43	11.13	1.09
1–20.....	117.22	48.72	4.85	103.81	49.66	4.89

Trials	Bell-Shock-Wrong (III)			Bell-Shock-Right (IV)		
	Mean	S.D.	P.E.$_m$	Mean	S.D.	P.E.$_m$
1– 5.....	59.76	11.33	1.31	50.50	12.24	1.42
6–10.....	37.91	15.08	1.75	30.18	15.88	1.84
11–15.....	21.50	16.14	1.87	16.29	14.73	1.70
16–20.....	9.59	11.99	1.39	8.47	11.19	1.29
1–20.....	128.82	45.99	5.32	105.71	48.44	5.60

TABLE 4

COMPARISON OF GROUPS WITH ONE ANOTHER

Diff. = difference between means of groups; P.E.$_{diff.}$ = probable error of difference; C.R. = critical ratio or difference divided by P.E.$_{diff.}$; P = chances in 100 of such a difference being a true difference and not one due to chance.

Trials	Bell-Wrong (I) minus Bell-Right (II)				Bell-Shock-Wrong (III) minus Bell-Shock-Right (IV)			
	Diff.	P.E.$_{diff.}$	C.R.	P	Diff.	P.E.$_{diff.}$	C.R.	P
1– 5..	5.79	1.86	3.11	96	9.17	1.93	4.75	100
6–10..	5.87	2.40	2.44	90	7.73	2.53	3.05	96
11–15..	1.53	2.02	.76	39	5.21	2.53	2.06	83
16–20..	−.06	1.53	−.04	3	1.12	1.90	.59	31
1–20..	13.41	6.88	1.95	81	23.11	7.73	2.99	96

Trials	Bell-Shock-Wrong (III) minus Bell-Wrong (I)				Bell-Shock-Right (IV) minus Bell-Right (II)			
	Diff.	P.E.$_{diff.}$	C.R.	P	Diff.	P.E.$_{diff.}$	C.R.	P
1– 5..	3.35	1.77	1.89	80	−.03	2.01	−.01	00
6–10..	2.74	2.49	1.10	54	.88	2.45	.36	19
11–15..	4.31	2.37	1.82	78	.63	2.20	.29	16
16–20..	1.22	1.75	.69	36	.04	1.70	.02	1
1–20..	11.60	7.19	1.61	72	1.90	7.43	.25	13

probably significant save that between the simple Bell-Right curve and the Bell-Shock-Right curve. Bell on the wrong holes makes it more difficult to learn the total task than bell on the right holes. And, adding shock to the wrong holes in addition to bell makes the total task even more difficult. Bell on the right holes and bell plus shock on the right holes make the total task easier—but there is no very clear difference between bell alone and bell plus shock.

Before, however, attempting any final interpretation of the above results we must consider the results of Group I*a*. For it must be noted that where the Bell-Wrong and the Bell-Shock-Wrong groups learned one pattern of holes (the unstarred series in Fig. 1), the Bell-Right and the Bell-Shock-Right learned the other pattern of holes (the starred series in Fig. 1). And it might be that the obtained differences, that is, the general superiority of the Bell-Right and the Bell-Shock-Right groups over the other two, was merely due to the one pattern being more difficult than the other. It was to test this out that Group I*a* was run. For this group the holes unstarred (Fig. 1) were connected with the buzzer. In other words, Group I*a* like Group I were a Bell-Wrong group but actually they were required to select the opposite set of holes from those of Group I—that is, to select the same absolute holes as those selected by Group II and Group IV, the Bell-Right and the Bell-Shock-Right groups. Group II and Group IV should therefore also be compared with this Group I*a* to make sure that the differences we found between them and Group I were not meretricious and due merely to differences in the pattern of holes as such. In Fig. 3 we show the cumulative curves for Group I, Group I*a*, and Groups II and IV, and in Tables 5 and 6 we show the statistical values for this group and for this group compared with Group I and Groups II and IV.

It appears from these tables and from the curves in Fig. 3 that Group I*a* was somewhat better than Group I but not as much better as were Group II or Group IV. We conclude that the Bell-Right and Bell-Shock-Right were in all probability really better than Bell-Wrong and really even still better than Bell-Shock-Wrong.

Before turning to our final conclusions and interpretations we may note the high reliabilities of the individual differences within each of the four main groups. These reliabilities are presented in Table 7. The coefficients were computed by correlating the errors on the first five odd trials with the errors

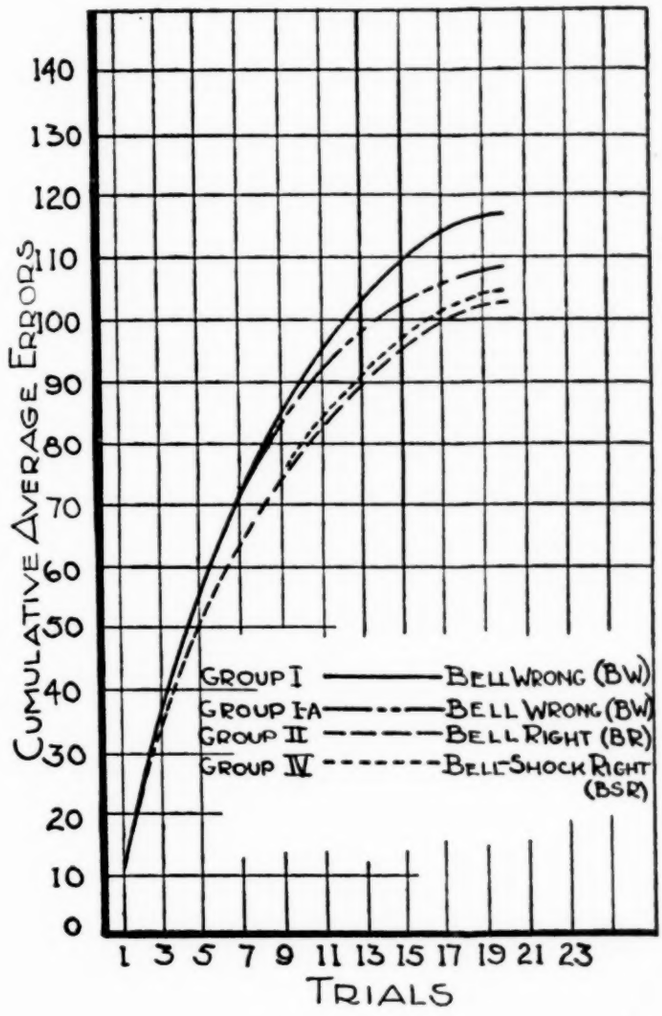


FIG. 3. Trials.

on the first five even trials, and estimating the reliability of all ten trials by means of the Spearman-Brown formula. The coefficients are all over .90, indicating that whatever ability is measured by performance on this task is being measured consistently throughout the ten trials.

Interpretations.—We turn now to the question of interpretations.

First, let us seek to explain the effects of just the bell

TABLE 5

MEAN, STANDARD DEVIATION, AND PROBABLE ERROR OF THE MEAN FOR THE BELL-WRONG GROUP (I*a*) WHICH LEARNED THE PATTERN OPPOSITE THAT OF THE BELL-WRONG GROUP (I)

Trials	Bell-Wrong (I*a*)		
	Mean	S.D.	P.E.$_m$
1– 5	55.24	11.42	1.54
6–10	33.08	13.81	1.86
11–15	14.60	15.81	2.13
16–20	5.80	11.69	1.58
1–20	108.64	45.84	6.18

TABLE 6

COMPARISON OF GROUP I*a* WITH EACH OF THE OTHERS
(See Table 3 for explanation of column headings)

Trials	Bell-Wrong (I) minus Bell-Wrong (I*a*)				Bell-Wrong (I*a*) minus Bell-Right (II)				Bell-Wrong (I*a*) minus Bell-Shock-Right (IV)			
	Diff.	P.E.$_{diff.}$	C.R.	P	Diff.	P.E.$_{diff.}$	C.R.	P	Diff.	P.E.$_{diff.}$	C.R.	P
1– 5.	1.17	1.95	.60	31	4.62	2.10	2.20	86	4.74	2.09	2.27	88
6–10.	2.09	2.57	.81	41	3.78	2.47	1.53	69	2.40	2.62	1.11	54
11–15.	2.59	2.58	1.00	50	−1.06	2.54	− .42	21	−1.69	2.73	− .62	31
16–20.	2.57	1.91	1.35	64	−2.63	1.92	−1.37	64	−2.67	2.04	−1.31	62
1–20.	8.58	7.86	1.09	54	4.83	7.88	.61	31	2.93	8.34	.35	19

TABLE 7

RELIABILITY COEFFICIENTS FOR THE FOUR MAIN GROUPS

Comparable Scores Gotten by Taking First Five Odd Trials Against First Five Even Trials

Groups	Reliability Coefficients		
	n	Correlation Between First 5 Odd and First 5 Even Trials	Reliability Coefficient of First 10 Trials (Brown-Spearman Formula)
Bell-Wrong	46	.906	.951
Bell-Right	47	.941	.970
Bell-Shock-Wrong	34	.854	.921
Bell-Shock-Right	32	.901	.948

alone. For it appears quite definitely that the bell simply by itself when it occurs on the right responses and is the signal for those responses makes learning easier than when it occurs on, and is the signal for, the wrong responses. But bell alone

seems hardly a matter of either reward or punishment. Rather it seems to be a *signal* or *emphasizer*. It *signals* to the subject that the response just made is right (or vice versa that it is wrong) and it thereby also *emphasizes* the one hole or the other. And, turning now to the actual outcome, it appears that when this signal or emphasizer is on the correct hole it is easier for the subject to learn than when it is on the incorrect hole. And, if we may envisage learning as consisting in the building up of little subordinate gestalts in which the one hole of each pair is to be set off against the other hole of the pair,[4] this is perhaps just what we might expect. For it would seem quite reasonable that if the positive *emphasis*, or *accent*, is on the to-be-selected hole, learning should be easier than when this positive emphasis or accent is on the to-be-avoided hole. Indeed it would seem that the present finding is really in line with that obtained by Thorndike (**2**), p. 45, when he discovered that it aided learning more to tell a subject when he was right than to tell him when he was wrong. It does not seem to us, in short, reasonable to interpret either bell or the mere words 'right' and 'wrong' as rewards or punishments. The true rewards and punishments—satisfiers and annoyers—are the ultimate 'getting-to' or 'not-getting-to' the end of the task and the getting there successfully. The effect of the bell (or of the word 'right' or 'wrong') is merely to emphasize. And our finding is that an emphasis upon success is more helpful than an emphasis on lack of success. The emphasis upon the latter would seem to act, perhaps, as a baleful fascination which attracts the performer and which makes the avoiding of the response in question the more difficult.

In the second place, we must seek to consider the results when the shock was added to the bell. Now the shock was something of a punishment in and of itself. It acted as a *motivator*. And, if this were all, then the group who received the shock upon the wrong responses (*i.e.*, *Group III*), the Bell-Shock-Wrong group, should have learned faster than the corresponding group who did not receive shock (*i.e.*, Group I),

[4] For an elaboration of the present argument see Tolman (**3**), Chapters XXII and XXIII especially, 344–346 and 362–367.

the simple Bell-Wrong group, since learning in Group III would be in accord not only with the motive for general success but also in accord with that for an avoidance of shocks. Actually, however, the Bell-Shock-Wrong group, Group III, learned more slowly than the Bell-Wrong group, Group I.

Now one way that we can explain this would seem to be to point out that shock is also an additional emphasizer and that, as such an additional emphasizer on the wrong responses, it would work against rather than for more rapid learning. Its emphasizing effect here tended to counteract its effect as something in and of itself to be avoided.

Finally, however, let us consider the effect of the shock upon the right responses. As emphasis the shock on the right responses should according to our above reasoning have aided learning. As punishment in and of itself it should neither have aided nor hindered since, whether or not the wrong (and in this case unpunished) response was given first, the right and punished response always had to be given afterwards. In other words the avoidance of learning did not, in this set-up, diminish the amount of shock in each trial. Indeed considering all the trials together it would appear that the fewer the trials the smaller the total amount of shock so that, if anything, the shock for this group also should have caused faster learning. Actually, however, we see that the Bell-Shock-Right group did not do consistently better than the Bell-Right. Neither, however, did it do appreciably worse.

Still another hypothesis seems to be needed. It has several times been reported by other investigators that electric shock as such has a *disrupting* or distracting effect (1, 4). As such it would tend to hinder learning wherever it is introduced. Introduced after the wrong responses it would tend to cause distraction and to prevent them being remembered as wrong, and introduced after the right responses it would tend similarly to cause distraction and to interfere with their being remembered as right. In short, on this hypothesis, this disruptive effect of the shock would, when added to the wrong responses, work in the same direction as its emphasizing effect and further hinder learning. And this would help to explain

the decidedly greater poorness of the Bell-Shock-Wrong group compared with the Bell-Wrong group. And when added to the right responses this disruption effect of the shock would work against its emphasizing effect. And this would explain why the results for the Bell-Shock-Right group are not clearly worse or better than those for the Bell-Right group.

Conclusion.—In conclusion it may be well to note that actually we have been led to suggest three laws of learning as a substitute for and correction of Thorndike's law of effect. These would be a Law of Motivation, a Law of Emphasis and a Law of Disruption.

By a Law of Motivation we would mean the assumption that the ultimate reason for learning seems to be the obtaining of final successes or of final failures. (And, insofar as the avoidance of shocks is in and of itself a potent goal, then the introduction of shock to the wrong responses will favor learning.)

By a Law of Emphasis we would mean the assumption that in a trial and error situation learning consists in building up little patterns or gestalts in which the correct response is set over against the incorrect one or ones and the further assumption that an accent or emphasis (such as bell or shock) upon the correct responses will tend to favor learning whereas an accent or emphasis upon the wrong responses will tend to hinder learning.

By a Law of Disruption we would mean the assumption that any relatively violent negative emotional stimulus (such as an electric shock) coming in immediate sequence upon either the right or the wrong responses will, as such, tend to disrupt learning.

Finally we may indicate once again the application of all three of these laws to our actual findings.

The Bell-Wrong group should be worse than the Bell-Right group by virtue of the Law of Emphasis and this was the actual outcome.

The Bell-Shock-Wrong group when compared with the Bell-Wrong group should be worse than the latter by virtue of both the Law of Emphasis and the Law of Disruption. It

should be better than the latter by virtue of the Law of Motivation. Actually it was worse.

The Bell-Shock-Right group should be better than the Bell-Right group by virtue of the Law of Emphasis and possibly also a little bit better by virtue of the Law of Motivation. It should, however, be worse than the Bell-Right group by virtue of the Law of Disruption. Actually it was about the same as the simple Bell-Right group.

One last word—the present experiment and this final discussion are presented for their suggestive value and not in any belief that they are definite or final. Our three 'laws' are mere tentative hypotheses to be further investigated.

References

1. Rexroad, C. N. Administering electric shock for inaccuracy in continuous multiple choice reactions. *J. Exper. Psychol.*, 1926, **9**, 1–19.
2. Thorndike, E. L. *Human Learning.* New York: The Century Co., 1931.
3. Tolman, E. C. *Purposive Behavior in Animals and Men.* New York: The Century Co., 1932.
4. Vaughn, J. and Diserns, C. M. The relative effects of various intensities of punishment on learning and efficiency. *J. Comp. Psychol.*, 1930, **10**, 55–66.

(Manuscript received April 15, 1932)

THE PREDICTION OF VOCATIONAL APTITUDE AND SUCCESS FROM PHOTOGRAPHS [1]

BY MORRIS S. VITELES AND KINSLEY R. SMITH
University of Pennsylvania

Objectives

The objectives of the study described in this report were:

(1) To investigate the accuracy of judgments on vocational aptitude and success or failure made by experienced personnel workers, from an examination of photographs.

(2) To determine the degree of certainty with which such judgments are made.

(3) To analyze the specific physical and other signs employed when making such judgments.

(4) To compare the judgments and degree of certainty of experienced personnel workers with those of observers inexperienced in the selection and management of workers.

Subjects and Procedure

(*A*) *The Subjects*

The procedure employed in this investigation followed that used by Landis and Phelps in a study published in 1928.[2] The only important difference is in the choice of subjects. The subjects of the investigation by Landis and Phelps were college students. In the present study judgments were made by twenty-four members of a personnel association of a large Eastern city. The majority of the subjects are employment managers or interviewers in industrial plants. A few are employed in placement work in public and private employ-

[1] Paper read at the annual meeting of the American Psychological Association Toronto, Sept. 12, 1931.

[2] Landis, C. and Phelps L. W., 'The Prediction from Photographs of Success and of Vocational Aptitude,' this Journal, 1928, 11, 313–24. See this article for review of earlier studies.

ment agencies. Practically all have had occasion to use photographs in making judgment on personality in connection with employment or placement. Nearly every member of the group is acquainted with one or another physiognomic system of character analysis. Approximately ten of this group of personnel workers have received formal training in physiognomic systems of character analysis, and three or four make strict applications of such systems in their daily work.

In the opinion of the authors, the most significant feature of this experiment is the opportunity which it allows for a direct comparison of the judgments of observers experienced in personnel work with those of subjects untrained in the art of selecting and placing workers, and without experience in this field.

(*B*) *The Material*

The material employed in this experiment was obtained through the courtesy of Dr. Carney Landis. It consists of two photographs of five *successful* and five *unsuccessful* men in each of the fields of *law*, *medicine*, *education* and *engineering*, making a total of twenty successful and twenty unsuccessful men. Landis and Phelps describe as follows the method and criteria employed in the selection of the photographs.

"An alumni publication of a large American university was secured. This publication was issued on the 25th anniversary of the graduation of a class of 850 men and it contained an autobiographical sketch of each man together with his photograph *at the time of graduation* and at his *25th anniversary of graduation*. For the purposes of this study, all autobiographies of men who had gone into law, medicine, education and engineering were read and assigned a rating from 1 to 10 on the basis of their apparent wordly success. For instance, a lawyer who had risen to be one of the chief attorneys for the Standard Oil Company was judged relatively successful, while one who was acting as clerk in a large law office was relatively unsuccessful. A man who had taught for 25 years in the common schools was judged relatively unsuccessful, while one who had risen to the superintendency of schools in a large city was judged relatively successful. In this study it must be borne in mind that success or non-success relates only to the vocation and is always a relative term—relative to the performance of other members of his own college class.

"From the ratings thus assigned the five most successful and the five least successful in each of the four groups, law, medicine, education and engineering were taken. The published portraits were rephotographed and mounted on separate cards."

Landis and Phelps point out the advantage of this method of selecting photographs for use in an experimental investigation of predicting success or failure from photographs. The selection was made not on the basis of the appearance of success or failure, but on the basis of an autobiography which gives a clear picture of how successful each individual has been

in the period of 25 years subsequent to graduation from college. In this respect, as Landis and Phelps point out, the investigation differs from earlier studies, such as those of Gaskill, Fenton and Porter [3] and of Pintner,[4] in which photographs were not selected on the basis of complete factual evidence with respect to the 'strength' of the trait under consideration.

(C) *Procedure*

The photographs were divided into two groups,—one of 'younger' men, including the photographs taken at the time of graduation, and one of 'older' men, including the photographs taken at the 25th anniversary of graduation. The forty photographs in each group,—twenty of successful and twenty of unsuccessful men,—were scattered in random order with respect to success and with respect to actual vocation. The photographs in each group were projected one at a time upon a screen. The photographs of the 'successful' and 'unsuccessful' or 'failure' men, taken at the time they were college seniors, were presented first. Shortly afterwards the pictures of the 'older' men were shown. However, the subjects were not informed that the two groups of pictures were of the same men.

Each subject was given two blanks on which to record his judgments. Directions on the prepared blanks were read aloud while the observers followed them from their individual copies. The blank used by the observers in recording judgments in the case of the younger men bore the following instructions:

"This experiment is designed to test your ability to predict from photographs the relative vocational success or failure of college seniors, and your ability to predict what profession, vocation, craft, or trade these men will follow.

"I will project singly on the screen the photographs of 40 young men. The photographs of these young men were taken when the men were seniors in college some twenty-five years ago. Certain of these pictures are of men who will be very successful in the professions which they follow. Others of the pictures are of men who will be relatively unsuccessful in their professions. Note then that there are two groups, one of very successful men and one of relatively unsuccessful men.

"Look at each picture in turn and then record your judgment in the appropriate column.

"In the column headed 'Success or Failure' place an 'S' or 'F.' 'S' signifies that you believe that the man will be relatively successful in whatever vocation he may follow. 'F' means that he will be relatively unsuccessful.

"In the next column record the 'Percent Certain' of this judgment. 0 percent means a hunch or no feeling of certainty whatsoever; 50 percent moderately certain; and 100 percent means positive conviction.

"In the column headed 'Vocation' enter the number from the card (Form C A 1) indicating profession, trade, etc., which you would recommend.

"In the column headed 'Basis of Judgment' enter the number from the card (Form C A 1) showing the reason best describing your basis of judgment."

---

[3] Gaskill, P. C., Fenton, N., and Porter, J. P., 'Judging the Intelligence of Boys from Their Photographs,' *J. Appl. Psychol.*, 1927, 11, 394–404.

[4] Pintner, R., 'Intelligence as Estimated from Photographs,' *Psychol. Rev.*, 1918, 25, 286–96.

Below the instructions were six columns headed as follows:

Picture Number	Success or Failure	Percent Certain	Vocation	Basis of Judgment	Remarks

There was a short intermission between the projection of the pictures of the 'younger' and of the 'older' groups. As indicated above, the observers were not told that the two groups of pictures were of the same men. The blank and the instructions used in obtaining the judgments on the photographs of the older men were largely the same as those employed in the case of the younger men, but the observers were instructed to make a judgment on whether the man *had been* successful, and to indicate the profession which he *had followed*. The first two paragraphs on the blank used in obtaining judgments on photographs of the older men read as follows:

"This section of the experiment is designed to test your ability to judge experimentally from photographs what vocation, profession, or trade men have followed for 25 years, and whether they have been very successful or relatively unsuccessful.

"I will project singly on the screen the photographs of 40 men. These men had been graduated from college some 25 years at the time the photographs were taken. Certain of these pictures are of men who have been very successful in the profession which they have followed. Others of the pictures are of men who have been relatively unsuccessful in their professions. Note then that there are two groups, one of very successful men and one of relatively unsuccessful men."

Each subject was handed a sheet (Form C A 1) showing a classification of possible vocations, and a list of 'reasons for judgments.' A section of the list of 'possible vocations' and the entire list of 'reasons for judgments' are reproduced below.

Form C A 1

POSSIBLE VOCATIONS

1. Law
   - 1A General practice
   - 1B Politics
2. Business
   - 2A Wholesale merchandising
   - 2B Retail merchandising
   - 2C Finance
   - 2C 1 Banking
   - 2C 2 Brokerage
   - 2C 3 Insurance
3. Education
   - 3A College instruction
   - 3B College president
4. Intellectual vocations other than teaching
   - 4A Clergyman
   - 4B Scientist
   - 4C Inventor
   - 4D Author
   - 4E Poet

REASONS FOR JUDGMENTS

1. Shape of face
2. Appearance of eyes
3. Length of nose
4. Shape or size of mouth
5. Width of forehead
6. Height of forehead
7. Shape or size of ears
8. Condition (neatness, etc.)
9. Facial expression
10. Position of head

Form C A 1—*continued*

Possible Vocations	Reasons for Judgments
5. Engineering	11. Similarity to some acquaintance
5A Electrical	
5B Mechanical	12. Observation of persons of similar appearance
6. Medicine	
6A General practice	
6B Surgeon	13. General impression
7. Agriculture	
7A Farming	14. Hunch
7B Forestry	
7C Ranching	15. Snap Judgment
7D Country gentleman	
8. Gentlemen of leisure	16. Unclassified

Summary of Findings

1. The experimental evidence obtained in this investigation is in contradiction to the claims of physiognomists, and to the popular belief of personnel workers, that vocational aptitude and success can be judged from photographs.

2. The prediction and judgment of vocational aptitude and of success or failure from photographs by personnel workers are no better than chance.

3. The judgment and prediction of experienced personnel workers are no more accurate than those of untrained observers (college undergraduates).

4. In spite of the fact that the accuracy of prediction and of judgment of personnel workers is no better than that of college students, personnel workers are more certain of the accuracy of their judgments.

5. Personnel workers seem to have a bias toward success. A higher percentage of 'success' judgments reveals a degree of optimism with respect to vocational success greater than that of college students.

6. The judgments of personnel workers, as the judgments of college students, do not seem to be based upon any single system of character analysis. They depend largely upon *general impression*, *facial expression*, and *resemblance to an acquaintance*.

The findings upon which these conclusions are based are embodied in the accompanying tables and charts.

Analysis of Findings

(1) *The Ability of Personnel Workers to Judge Vocational Success from Photographs of 'Older' Men*

Table I and Chart III present evidence on whether personnel workers can judge vocational success from the examination of photographs of men in professional work who have been out of the University for twenty-five years.

The twenty-four personnel workers recorded a total of 475 judgments in the case of the 'most successful' older group, and a total of 478 judgments with respect to the vocational success of the 'least successful' or 'failure' group. The table shows that 82.7 percent of the judgments of 'successful' men are 'success' or correct judgments. The percentage of wrong or 'failure' judgments in the case of the older 'successful' group amounts to only 17.3 percent. On the other hand 76.8 percent of the judgments of the 'least successful' or 'failure' group are also 'success' judgments. In other words, only 23.2 percent of the judgments concerning the degree of success of the 'least successful' or 'failure' group are correct.

If we combine the judgments from which these percentages are derived it is found that in the case of the personnel workers 52.8 percent of the judgments are correct (Chart III). In other words, the percentage of correct judgments is very slightly better than chance. If the members of this group had tossed coins instead of attempting to make a deliberate judgment on the relative success of the men whose photographs were shown, the results would have been just as accurate as those obtained through deliberate judgment.

If the photographs of the 'older' group are considered individually the invalidity of judgments on 'success' or 'failure' is again evidenced. Of the 20 'older' men who were actually 'successful' 19 were judged 'successful' by a large majority of the observers. Of the 20 who were actually 'least successful' or failures, 18 were judged 'successful' by the majority of personnel workers. It is interesting to note that in the case of four photographs of the 'older' group there is unanimous agreement with respect to degree of success. All

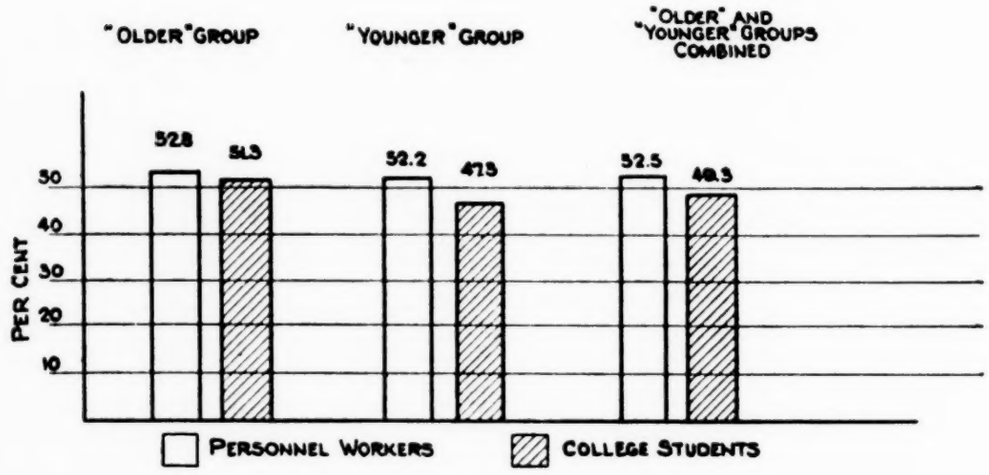
TABLE I

Success and Failure Judgments—Photographs of 'Older' Group

Successful Group									Failure Group								
Picture	No S Jdg-mts.	No F Jdg-mts.	Per-cent S Jdg-mts.	Landis-Phelps	Per-cent F Jdg-mts.	Landis-Phelps	Av. Cert'y	Landis-Phelps	Picture	No S Jdg-mts.	No F Jdg-mts.	Per-cent S Jdg-mts.	Landis-Phelps	Per-cent F Jdg-mts.	Landis-Phelps	Av. Cert'y	Landis-Phelps
S1-O	19	5	79	*71*	21	*29*	52.6	*40.7*	F21-O	19	5	79	*35*	21	*65*	50.2	*36.7*
S2-O	23	1	96	*81*	4	*19*	64.5	*39.7*	F22-O	23	0	100	*67*	00	*33*	57.6	*44.4*
S3-O	17	7	71	*38*	29	*62*	51.7	*28.5*	F23-O	22	2	92	*77*	8	*23*	61.3	*32.3*
S4-O	23	1	96	*68*	4	*32*	49.3	*35.5*	F24-O	15	9	63	*42*	37	*58*	48.9	*32.0*
S5-O	20	4	83	*47*	17	*53*	56.1	*30.7*	F25-O	24	0	100	*71*	00	*29*	61.7	*28.0*
S6-O	23	1	96	*67*	4	*33*	61.9	*40.0*	F26-O	16	8	67	*36*	33	*64*	53.2	*28.3*
S7-O	20	4	83	*45*	17	*55*	59.5	*36.5*	F27-O	13	11	54	*26*	46	*74*	49.5	*41.2*
S8-O	19	4	83	*24*	17	*76*	53.0	*28.2*	F28-O	24	0	100	*51*	00	*49*	57.2	*32.7*
S9-O	23	1	96	*60*	4	*40*	61.5	*31.3*	F29-O	17	7	71	*25*	29	*75*	54.1	*35.2*
S10-O	22	2	92	*46*	8	*54*	54.8	*34.2*	F30-O	9	15	38	*20*	62	*80*	47.0	*38.3*
S11-O	21	3	88	*52*	12	*48*	52.8	*30.0*	F31-O	16	8	67	*40*	33	*60*	55.6	*39.7*
S12-O	24	0	100	*60*	0	*40*	73.4	*24.1*	F32-O	10	14	42	*40*	58	*60*	46.5	*34.4*
S13-O	21	3	88	*49*	12	*51*	56.7	*33.6*	F33-O	23	1	96	*60*	4	*40*	52.8	*42.0*
S14-O	17	7	71	*41*	29	*59*	53.7	*36.3*	F34-O	18	5	78	*60*	22	*40*	51.1	*36.3*
S15-O	17	6	74	*68*	26	*32*	60.2	*29.2*	F35-O	21	3	88	*55*	12	*45*	47.1	*31.9*
S16-O	16	8	67	*16*	33	*84*	47.7	*26.7*	F36-O	22	2	92	*46*	8	*54*	61.0	*35.0*
S17-O	22	1	96	*65*	4	*35*	60.0	*38.1*	F37-O	23	1	96	*83*	4	*17*	59.1	*43.6*
S18-O	15	8	65	*30*	35	*70*	56.8	*34.8*	F38-O	20	4	83	*44*	17	*56*	56.9	*34.7*
S19-O	10	14	42	*18*	58	*82*	43.9	*29.7*	F39-O	16	8	67	*46*	33	*54*	49.3	*36.5*
S20-O	21	2	91	*61*	9	*39*	49.7	*36.7*	F40-O	16	8	67	*30*	33	*70*	49.3	*27.7*
Av.	19.6	4.1	82.7	*50.4*	17.3	*49.6*	56.5	*34.2*		18.3	5.5	76.8	*47.8*	23.2	*52.2*	53.5	*35.5*

observers unite in making a 'success' judgment in the case of S12–0 who actually belongs in the 'successful' group. At the same time all observers unanimously place F22–0, F25–0 and F28–0 in the 'most successful' group, although these three actually belong in the 'least successful' group. In other words, in cases of individual photographs where unanimous agreement exists among the observers, the observers unanimously agree on wrong judgments in three cases out of four.

CHART Nº III
PERCENTAGE OF CORRECT JUDGMENTS
ON "SUCCESS" AND "FAILURE"



(2) *The Ability of Personnel Workers to Predict Vocational 'Success' from Photographs of 'Younger' Men*

In Table II and Chart III are embodied results of the section of the experiment dealing with the prediction of vocational 'success' or failure from an examination of the photographs of the 'younger' men.

The twenty-four personnel workers recorded a total of 472 judgments in the case of the 'most successful' younger group, and a total of 473 judgments with respect to the vocational 'success' of the 'least successful' or 'failure' group. Table II shows that 80.5 percent of the predictions in the case of the 'successful' younger men are 'success' or correct predictions. The percentage of wrong or 'failure' predictions in the case of the 'younger' 'most successful' group amounts to only

TABLE II

SUCCESS AND FAILURE JUDGMENTS—PHOTOGRAPHS OF 'YOUNGER' GROUP

Successful Group									Failure Group								
Picture	No S Jdg-mts.	No F Jdg-mts.	Per-cent S Jdg-mts.	Landis-Phelps	Per-cent F Jdg-mts.	Landis-Phelps	Av. Cert'y	Landis-Phelps	Picture	No S Jdg-mts.	No F Jdg-mts.	Per-cent S Jdg-mts.	Landis-Phelps	Per-cent F Jdg-mts.	Landis-Phelps	Av. Cert'y	Landis-Phelps
S1-Y	20	4	83	*46*	17	*54*	54.1	*20.7*	F21-Y	23	1	96	*67*	4	*33*	55.9	*28.2*
S2-Y	19	5	79	*45*	21	*55*	48.9	*23.0*	F22-Y	15	9	63	*62*	37	*38*	53.4	*34.5*
S3-Y	22	2	92	*61*	8	*39*	56.7	*32.7*	F23-Y	14	9	61	*44*	39	*56*	48.7	*34.8*
S4-Y	18	6	75	*32*	25	*68*	44.5	*30.7*	F24-Y	10	14	42	*25*	58	*75*	46.5	*25.0*
S5-Y	21	2	91	*55*	9	*45*	49.7	*30.0*	F25-Y	19	4	83	*83*	17	*17*	57.3	*30.5*
S6-Y	18	6	75	*58*	25	*42*	54.0	*33.0*	F26-Y	23	1	96	*74*	4	*26*	51.8	*31.1*
S7-Y	18	6	75	*41*	25	*59*	55.2	*33.8*	F27-Y	11	12	48	*20*	52	*80*	43.2	*25.2*
S8-Y	17	7	71	*29*	29	*71*	46.5	*25.5*	F28-Y	19	4	83	*47*	17	*53*	49.5	*31.2*
S9-Y	22	1	96	*46*	4	*54*	56.1	*32.4*	F29-Y	17	6	74	*35*	26	*65*	50.0	*25.5*
S10-Y	21	3	88	*48*	12	*52*	50.2	*30.7*	F30-Y	19	4	83	*56*	17	*44*	50.2	*28.7*
S11-Y	22	2	92	*82*	8	*18*	62.1	*32.2*	F31-Y	13	11	54	*33*	46	*67*	44.1	*26.7*
S12-Y	20	3	87	*32*	13	*68*	59.3	*32.0*	F32-Y	22	2	92	*71*	8	*29*	63.3	*32.0*
S13-Y	20	4	83	*49*	17	*51*	49.3	*33.0*	F33-Y	23	1	96	*57*	4	*43*	67.1	*30.4*
S14-Y	23	1	96	*65*	4	*35*	65.4	*33.8*	F34-Y	19	4	83	*30*	17	*70*	58.8	*30.0*
S15-Y	16	6	73	*45*	27	*55*	49.0	*30.9*	F35-Y	17	7	71	*75*	29	*25*	43.1	*26.3*
S16-Y	22	1	96	*45*	4	*55*	68.9	*25.5*	F36-Y	18	6	75	*50*	25	*50*	40.9	*26.6*
S17-Y	18	6	75	*31*	25	*69*	48.4	*25.0*	F37-Y	19	5	79	*65*	21	*35*	51.0	*32.0*
S18-Y	16	7	70	*49*	30	*51*	40.0	*32.0*	F38-Y	18	6	75	*41*	25	*59*	50.4	*35.2*
S19-Y	14	10	58	*34*	42	*66*	47.0	*29.5*	F39-Y	17	7	71	*43*	29	*57*	55.7	*29.5*
S20-Y	13	10	57	*36*	43	*64*	53.2	*20.0*	F40-Y	23	1	96	*59*	4	*41*	60.4	*22.9*
Av.	19.0	4.6	80.5	*46.6*	19.5	*53.4*	53.0	*29.3*		17.9	5.7	75.9	*51.9*	24.1	*48.1*	51.9	*29.3*

19.5 percent. On the other hand, 75.9 percent of the predictions of the 'least successful' or 'failure' group are also 'success' predictions. If figures from which these percentages are derived are combined, it is found that 52.2 percent of the predictions are correct. (Chart III.) In other words, the percentage of correct predictions on 'success' and 'failure' in the case of the 'younger' men is no better than chance.

An examination of the judgments on individual photographs of the 'younger' group in Table II shows again a somewhat uniform agreement with respect to classification of the photographs. However, in the 'younger' group there is no photograph in the case of which the personnel workers agree unanimously with respect to the classification on degree of 'success.'

(3) *Ability of Personnel Workers to Predict and Judge Vocational Aptitude*

A review of the judgments on the actual vocations of the 'older' men and of the recommendations with respect to vocations to be followed by the 'younger' men shows little agreement among the personnel workers. There is a wide range in the vocations assigned to each photograph. In the case of only twelve photographs do five or more judges agree on the vocation to be assigned to a man whose photograph is shown. In Table III are shown the photographs in the case of which

TABLE III

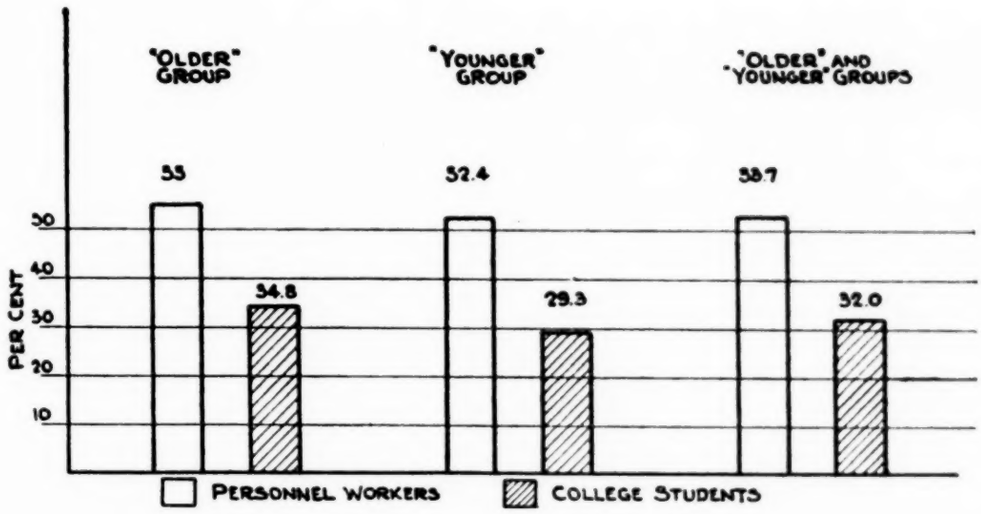
Picture	Assigned Occupation	Jdg-mts.	No. S	No. F	Actual Occup'n	Actual Standing
'Older' Men						
S 10-O	Clergyman	8	8	0	Engineer	S
S 14-O	Politics	8	5	3	Law	S
F 21-O	Civil Engineer	6	6	0	Educator	F
F 25-O	Banking	6	6	0	Educator	F
F 31-O	College Instructor	5	4	1	Law	F
F 36-O	Medical Specialist	5	5	0	Medicine	F
'Younger' Men						
S 2-Y	Clergyman	12	10	2	Educator	S
S 3-Y	Mech. Engineer	5	5	0	Educator	S
S 17-Y	Salesman	5	4	1	Medicine	S
F 29-Y	Clergyman	6	6	0	Law	F
F 34-Y	Musician	8	7	1	Engineer	F
F 40-Y	Clergyman	6	6	0	Educator	F

such agreement occurs. It is significant to note that in these twelve cases of comparative agreement there is only one, F36–O, in which the correct occupation is assigned. Eight personnel workers agree that S10–O is a clergyman. Actually he is a successful engineer. Twelve personnel workers agree that S2–Y, a successful educator, is a clergyman. Similar discrepancies are found in the remaining cases.

(4) *The Feeling of Certainty of Personnel Workers*

The degree of certainty on the part of personnel workers in making judgments on success and failure is shown in Chart IV. In view of the fact that the judgments made by the

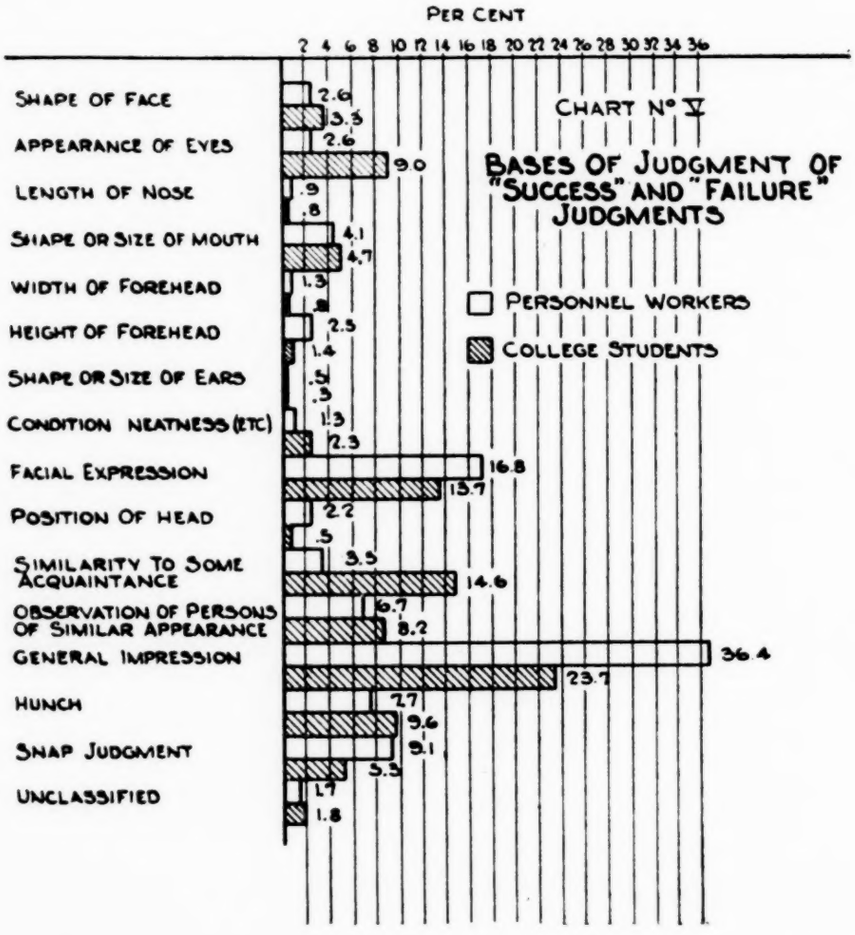
CHART Nº IV
CERTAINTY OF JUDGMENT OF "SUCCESS" AND "FAILURE" JUDGMENTS



personnel workers are actually no better than chance, an average feeling of certainty of 53.7 percent from all photographs can be described as extremely high. It is interesting to note that the range of average feeling of certainty for individual photographs when judged by the personnel group is from 40 percent to 73 percent. There is no significant difference in the feeling of certainty in predicting success and failure for the 'younger' group as compared with the feeling of certainty when passing judgment on the 'older' group.

(5) *The Physical Bases of Prediction and Judgment of Vocational Aptitude and Success and Failure*

In Chart V is shown the frequency with which each basis of judgment is employed by personnel workers in recording their predictions and judgments on the success and failure and the vocational aptitude of both the 'older' and 'younger' men. Personnel workers rely upon *general impression* and *facial expression* in the case of 53.2 percent of their judgments. An analysis of the data shows that they rely upon these two cri-



teria in the case of 61 percent of their judgments of 'older' men and in the case of 44 percent of their predictions regarding 'younger' men. It is significant to note that in few instances are the specific criteria of character analysis systems employed in arriving at a decision of success and failure. This is particularly significant in view of the acquaintance of this group of personnel workers with the most frequently employed sys-

tems of character analysis, and of the confidence of a number of them in their validity.

(6) *Comparison of Personnel Workers (Trained Observers) and College Students (Untrained Observers)*

In all of the tables and charts included with this report, the results of personnel workers are compared with those of untrained observers (college students) employed in the experiment by Landis and Phelps. In the tables, results of the college under-graduates used as subjects by Landis and Phelps are shown in italics. The cross hatched columns in the charts also show the results of the college undergraduates.[5]

An examination of Chart III shows that 52.8 percent of the judgments of personnel workers with respect to 'success' and 'failure' of the 'older' group are correct as compared with 51.3 percent [6] in the case of college students.

With respect to 'success' and 'failure' of the 'younger' group 52.2 percent of the judgments of personnel workers are correct as compared with 47.3 percent of the judgments of college students. In other words, the judgment of personnel workers with respect to 'success' and 'failure' of men whose photographs were shown, is very slightly better than that of college undergraduates, and like that of college students, is close to 'chance.'

An examination of Tables I, II and Chart III shows a rather interesting difference with respect to the tendency of judgment. Although there is no significant difference in the percentage of correct judgments between personnel workers and college students, personnel workers show a marked tendency to judge the photographs as 'successful.' Thus, in the case of the 'older' group 82.7 percent of the judgments were

[5] The figures for college students in Tables I, II, and Chart III are taken from a control experiment by Landis and Phelps, involving 93 students, in which directions had been changed so as to counterbalance a bias toward 'success' judgments which appeared in an earlier investigation.

The figures for college students in the remaining tables and charts are cited from the original experiment by Landis and Phelps, involving 20 subjects.

[6] This figure was obtained by averaging the average percent 'S' judgments and the average percent 'F' judgments given by Landis and Phelps. Inasmuch as the number of cases remains constant in the figures cited by Landis and Phelps, the figures on percent correct obtained in this way and shown on Chart III are valid.

'success' judgments. At the same time 76.8 percent of the judgments with respect to failures were 'success' judgments. The same tendency is revealed in the case of the 'younger' group, the judgments of 'success' being 80.5 percent in the case of the 'successful' group and the prediction of 'success' 75.9 percent in the case of the 'least successful' or 'failure' group. The undergraduates, on the other hand, furnish approximately the same proportion of 'success' and 'failure' judgments in the case of 'successful' and 'failure' groups both for 'older' and 'younger' men.[7]

When the trained and untrained observers are compared with respect to their ability to judge vocational aptitude, or what vocation a man has followed or will follow for twenty-five years, the personnel workers exhibit no more ability than do the college students. Landis and Phelps, as the authors of the present study, found a wide range of vocations assigned to each photograph. It seems that the personnel workers agree less among themselves in determining the vocations than do the college students, for, in spite of the fact that there were twenty-four personnel judges and only twenty student observers, the results of the former reveal only twelve photographs in the case of which the same vocation is assigned by five or more judges, while the results of the latter show fourteen such subjects. However, the inaccuracy in choice of vocation is about as great in both groups.

Landis and Phelps found the average feeling of certainty of college undergraduates in making ratings on 'success' and' failure' to be 32 percent when the 'older' and 'younger' groups are combined. (Chart IV.) For the combined groups the average feeling of certainty of the personnel worker is 53.7 percent, or approximately 1.6 times that of the college undergrad-

[7] Landis and Phelps found a high degree of optimism in the judgments of the photographs on the part of college undergraduates in their original experiment. The authors believed this to be due to the wording of the instructions. In a control experiment with revised instructions this tendency disappeared. The directions used in the present experiment were derived from the control experiment by Landis and Phelps. In spite of this fact, the percentage of 'success' judgments and predictions of personnel workers are actually in excess of those in the original experiment by Landis and Phelps, in which the instructions created a bias in the direction of 'success' judgments.

uates. The chart shows that college undergraduates tend to be more certain of their judgments of 'older' men than in predicting 'success' and 'failure' of 'younger' men. This tendency is also present in the case of personnel workers.

In Chart V is shown a comparison of the basis of the judgments employed by undergraduates and personnel workers. One of the most marked differences between personnel workers and college students is in *general impression.* The former depend upon general impression in the case of 36.4 per cent of the judgments, whereas college undergraduates depend upon this in the case of only 23.7 percent of their judgments. Another marked difference is to be found in *similarity to some acquaintance.* College undergraduates employ this criterion about four times as often as do personnel workers. Personnel workers do not tend to use the strictly physiognomic criteria of character analysis systems more frequently than do college undergraduates.

General Conclusions

In summarizing the results of their experiment, Landis and Phelps point out that their study "lends absolutely no evidence toward the confirmation of the claims of the physiognomists. It offers no evidence which would indicate that the inclusion of a photograph in a letter of application or on a personnel sheet is of any positive value." This earlier experiment was open to the criticism that its subjects were untrained college undergraduates whose experience in judging men was too restricted to permit an adequate judgment, regardless of the criterion employed. The present experiment shows that personnel workers, in spite of their experience in handling men, can use the photograph to no better advantage than can college students. These findings are of particular significance in view of the very widespread use of photographs, not primarily in industrial plants, but in selection by college authorities and in placement by employment agencies, of men and women in teaching, research work, and in other forms of professional work.

(Manuscript received October 1, 1931)

THE FUNCTIONAL SIGNIFICANCE OF IMAGERY DIFFERENCES [1]

BY FRANK C. DAVIS
University of California at Los Angeles

In 1914 Watson (**29**) 'dismissed' the image from psychology, justifying his stand by pointing to an alleged lack of correlation between the verbal reports of Fernald's (**8**) subjects and their objective performances in the 'imagery' tests devised by her. I have elsewhere (**4**) called attention to the fact that, had Watson been interested in making out a case for rather than against the image, he could have found numerous passages in Fernald's monograph indicative of the *presence* of correlation between subjects' reports and test scores, rather than the absence of correlation which he reported. Quite aside from this question of fact, however, it is clear that sweeping generalizations with respect to the objective significance of differences in imagery reports should not be based on results from a very small number of highly selected subjects. (Fernald's group comprised from five to eleven subjects, all of them graduate students or instructors in psychology.) The present investigation was planned to include a sufficiently large number of subjects to make conclusions therefrom of some statistical significance.

Statement of the Problem

The purpose of the present investigation may be stated as follows: To determine to what extent differences in subjects' reports as to their use of imagery in the performance of given tests are related to objective differences in test scores.

[1] The present report represents an abridgment of a thesis submitted in partial fulfillment of the requirements for the Ph.D. degree at the University of California. The investigation here reported was begun in 1924; it was discontinued when I left the university to take up other work, and was resumed when I returned to the university in 1928. I wish to acknowledge my indebtedness to Dr. Warner Brown, under whose direction this experiment was carried out, and to Dr. E. C. Tolman, who first stimulated my interest in the question of the objective significance of differences in imagery reports.

Subjects

The subjects were the students in the class in Elementary General Psychology. The majority of them were sophomores or juniors; a few were seniors or graduate students. The tests decided upon were administered as group tests to all the laboratory sections during the week when 'Imagery' was the topic under discussion in the class lectures. The number of subjects per section varied between fifteen and thirty. Each section gave fifty minutes to the experiment.

Method of Report Used by Subjects, and Preliminary Procedure

Before the administration of the tests was begun, the first ten minutes at each section meeting was given over to a brief recapitulation of the topic of imagery. The Galton 'breakfast table' type of procedure was gone through and the various possible modalities of images discussed. The names of the latter were written on the board and the subjects were told to use the first letters of each as abbreviations in their subsequent verbal reports.

Sometimes this recapitulation and discussion could not be arbitrarily terminated, promptly at the end of the ten minutes allotted, without an apparent sacrifice of rapport in the interests of a rigid time schedule. Since the establishment and maintenance of as high a degree of rapport as possible between subjects and experimenter was deemed of first importance, alternate tests were provided for use when the time became too short for the completion of all of the regular series of tests.

The subjects were always shown sample tests, and were instructed that at the end of each test they would be asked to write down the letter or letters representing the processes which they had used in recall or in recognition, as the case might be. If doubtful as to what imagery, if any, had been present they were told to put down a '?'. If sure that no imagery had played a part in their performance they were instructed to write 'none' at the end of the test or tests of which that seemed to them to be true. Where they were aware of more than one imaginal process in their performance

of a given task they were asked to underline the process or processes which seemed to them to be dominant, or if none of them seemed more dominant than the others to leave the letters un-underlined. (This was done only with the 1929 subjects. Many of the 1928 subjects *did* underline one or more processes, though uninstructed to do so. The distinction, seemingly, is difficult to make only in the case of those subjects whose imagery tendencies are 'balanced,' in the sense in which O'Brien (**18**) used that term. Most 'mixed' subjects report little hesitation concerning which process is for them dominant over the others.)

No attempt was made to differentiate between kinæsthetic imagery and kinæsthetic sensations—the subjects were told to put down 'K' if they were aware of movements, or images of movements, from lips, tongue, or larynx, or from other regions of the body, as playing a part in their recall of the material presented, or in the associations which came to them in the case of the 'Rhyming' and 'Childhood' association tests. (The tests used, and the manner of their selection, will be described in following sections.)

The tests proper were not begun until all questions had been answered, and until each subject apparently understood fully the various possible kinds of report. Then at the conclusion of the first test—that for Tonal Memory (see description following)—the experimenter always said: "Now state as accurately as possible how, in your opinion, you performed this test. (Slight pause.) Did you form a visual schema or pattern of the notes as you heard them the first time, and again the second time, and compare the second pattern with the first? Some people report that they do this sort of test that way. (Slight pause.) Or were you aware of movements, perhaps very slight movements, from your finger, or hand, or head, or eyes, or vocal cords, or from some other part of your body as the notes were played, and did you make your comparison of the two series of notes on the basis of those movements? Some people report that they do this sort of test in that way. (Slight pause.) Or did you concentrate on the sound of the notes in the first series, and then carry them in

your mind while you compared them with the notes of the second series? Some say they do it that way. Put down 'V' if a visual pattern was your method, 'K' if you were aware of movements and made your judgment on the basis of them, and 'A' if you made the comparison in terms of the sounds. If you used a combination of these or other processes put down the letters which stand for the processes you used. If one of them seemed to you to be dominant over the others, underline it. If you are sure none of these entered into your performance of the test put down 'none.' If you are uncertain, put down a question mark." [2]

At the end of each of the other tests the experimenter always said simply: "Now tell as accurately as you can how you did it."

Selection of Tests

In the matter of test selection I had the benefit of Fernald's extensive researches in this field. I made my initial selection largely from among her 'tests for more accurate analysis of type.' I eliminated some of these on the basis of my own experience with them, modified others for group administration, and added others of my devising. Among the tests which were eliminated were a 'spelling backwards' test, the

[2] Whatever element of suggestion entered into these remarks at the end of the Tonal Memory Test, I tried to weight in favor of reports of visual schema or kinæsthesis as against auditory imagery, by 'suggesting' the two former possibilities first. The extent to which subjects like these resist such suggestions, and the extent to which most of their reports are based on experiences the character of which seem to them rather unambiguous, was evident in their attitudes as they listened. A few sporadic nods of the head registered understanding and acquiescence in response to the 'visual pattern' and 'kinæsthesis' suggestions. The majority, however, merely waited, or registered downright negativism; which changed to affirmative head noddings as soon as the 'auditory memory' possibility was advanced. The suggestion that the fact that the stimuli were auditorily presented might account for the preponderance of auditory imagery reported as dominant in the case of this test may be countered by pointing to the results of the 'Rhyming' test, where although the stimuli were visually presented the reports were nevertheless preponderantly of auditory images as dominant, and by emphasizing once more the conclusions reached by Segal (23), Fernald (8), Feuchtwanger (9), O'Brien (18) and others, that dominant imagery tendencies are for many subjects more important than the physical characteristics of the stimulus in determining how the latter will be attended to. Note, in further corroboration of this, how in Table 2 the subjects' reports spread through various modalities and combinations of modalities no matter what the mode of presentation of the stimulus material.

reproduction of heard literary passages, the solving of physics problems and problems like the 'colored cube,' the 'straight-ahead' spelling of difficult English words, auditory span for numbers and letters, both forward and reversed, a 'title of book—name of author' association test, drawing from memory of a well-known face, free-hand drawing of the 'most-pleasing-rectangle,' ink-blot test for number and kind of associations. Of the above, I consider the spelling backwards test, the reproduction of heard literary passages, auditory span for numbers and letters, and perhaps the ink-blot test, promising tests for the diagnosis of imagery differences when used in the testing of individuals singly rather than in groups. The spelling backwards test I attempted to modify for use as a group test by making it a 'time-limit' test and scoring for number of errors and omissions. The 'time-limit' is difficult to enforce, however, since many individuals find themselves so stimulated by the novelty and the difficulties of such a task that they cannot, at command, cease their endeavors and turn at once to something else. The reproduction of heard literary passages took too much time if enough passages were used to furnish a measure of the test's reliability. The same was true of the ink-blot test. The solving of physics problems, in addition to consuming too much time, offered the further difficulty that many college students seem to find even the simplest problems in physics beyond them. 'Difficult English words,' for straight-ahead spelling, has certain drawbacks as a differential test of imagery tendencies, chiefly, it would seem from *a priori* considerations, on the score of unequal training, since methods of teaching spelling differ and since teachers themselves differ in their ability to 'put over' whatever methods are taught. However, when training may be considered to have been fairly constant there still is found great unevenness in spelling ability.[3] It was this observation that impelled me to consider such a test for inclusion as a possible imagery test, since it seemed conceivable that clear visual images in the case of both phonetic and nonphonetic spellings,

[3] De Voss (6) reports that of all the abilities measured by the Stanford Achievement Test battery, spelling is apparently most nearly 'independent' of the rest.

or clear auditory images in the case of phonetic spellings, might turn out to be important components of the 'spelling ability complex.' My early try-out of the test unfortunately uncovered a rather large proportion of subjects who failed so consistently on the words offered them that there seemed to be some danger of the disaffection engendered by their failure carrying over to the rest of the tests, so this was crossed off the list. But I feel that the question of a possible relationship between imagery tendencies and spelling ability does merit further investigation.

The elimination of the other tests which were tried was in every case on one of the following grounds: (1) scoring difficulties; (2) length of test necessary to assure anything like satisfactory reliability; (3) improbability, on the basis of a study of results from 'try-out' sections, that the test would show any relation to differences in imagery tendencies.

Description of Tests Selected

Results will be reported from the following tests: (1) The Seashore Tonal Memory test; (2) Memory for Simple Geometrical Figures; (3) Memory for Roman and Arabic Numerals; (4) Memory for Arabic Numerals; (5) Memory for Words Alike in Sound but not in Spelling; (6) Memory for Words Spelled Alike but Pronounced Differently; (7) Test of Rhyming Ability; (8) 'Childhood' Free Association Test.[4]

(1) *Tonal Memory.*—The Seashore test for tonal memory seems to offer an example of a test in which imagery may function in recognition. It forms one of the series of tests called by Seashore (21) 'Tests of Musical Ability.' These tests come in the form of phonograph records, one test to a record. Some of the tests have been reported as having low reliability, and their validity as tests diagnostic of 'musical ability' is in dispute. The 'Tonal Memory' test, however, yielded with the subjects of the present experiment a reliability coefficient of .90, and it has recently been reported by McCarthy (15) as giving a higher validity coefficient than any of the other Seashore tests. While I am not here concerned with the validity of the test as a measure of 'musical ability' but rather with its ability to discriminate individuals reporting different sorts of processes in their performance of it, nevertheless if it *is* to any degree a valid test of musical ability and if at the same time success in it should prove to be correlated with imagery reports of a given sort, then the question as to the effect of 'training' in imagery, upon musical ability, would evidently be in order.

In this test a series of tones varying in pitch is presented to the subject; then after a brief pause the same series, with the exception that the pitch of one of the tones is slightly raised or lowered, is repeated, and the subject is required to enter on a blank ruled for the purpose the number of the tone which was altered—whether, for example, when the series of tones presented numbered six, it was the first, second, third, fourth, fifth, or sixth which was different the second time from the first. The

[4] In addition to these, Army Alpha scores were available for correlational purposes.

total number of judgments to be made is fifty; the series of tones to be compared range in length from two to six, and ten judgments are made of each. The score used was simply the total number of right judgments.

(2) *Simple Geometrical Figures.*—In devising tests which involve the perception and reproduction of geometrical forms experimenters need to take note of the fact that the 'same,' from the purely objective standpoint, stimulus-pattern may be perceived quite differently by different subjects.[5] Where it is desired to place this variable as nearly as possible under control (where, for example, as in the present experiment, it is desired to make the test primarily a test of visual, or kinæsthetic, rather than of verbal memory) resort is had to a procedure akin to the nonsense-syllable technique of Ebbinghaus, and the visual patterns to be presented are made as nearly as possible 'meaningless.' In the present case, a large number of geometrical figures were shown, one after another, to the individuals comprising some of the 'try-out' sections mentioned previously, and these individuals were asked to write down all the verbal associations which each pattern 'set off.' The patterns to which it proved easiest to give verbal tags which seemed to aid materially in the subsequent reproductions were discarded, and those hardest to 'conceptualize' accurately during the short exposure periods were used in the later experiments. These selected patterns could conceivably be studied, and retained in memory, either visually or kinæsthetically or both in combination, but in the case of most of them, and for most if not all of our subjects, no simple verbal tag sufficed to assure accurate reproduction. The twelve figures chosen (see appendix) were shown to the subjects two at a time, and reproduction was called for immediately. The exposure time was five seconds for each pair of figures. Scoring was in terms of accuracy of reproduction. Each figure consisted of a curved line joined to two straight lines; if the reproduction was perfect (that is, if the position of the curved line with reference to the two straight lines, and the direction of all three lines, was like the original) it was scored '3,' if two thirds of it was reproduced correctly but the other segment was absent or wrong in direction the score given was '2,' if one segment was right but the other two in the wrong direction or omitted, the score was '1.' If a given pattern was all wrong, or omitted completely, '0' was recorded. Sample figures were shown, and the method of scoring was fully explained, before the test was given.

(3) *Roman and Arabic Numerals.*—This forms one of the group of tests labelled by Fernald 'tests for more accurate analysis of type.' Roman and Arabic numerals, in mixed order and arranged in Binet letter-square form, are presented to the subjects for a period of ten seconds. The subjects are instructed to reproduce them as accurately as possible. Scoring is in terms of number, and accuracy of placement, of

[5] Various investigators (Katz (13), Wulf (30), Messmer (16), Gibson (10), Allport (1), *e.g.*) have established this fact, which has led some of them to posit 'types' of observers—for example, 'central' and 'peripheral' types (Katz), 'comprehensive' and 'isolative' types (Wulf), 'subjective' and 'objective' types (Messmer). Whatever the merits of such attempts to classify subjects in rigid 'type' categories, the fact of individual differences in the perception of a given objective situation I believe is incontestable. O'Brien (18) reports that 'attentive selection,' exercised by his subjects 'among the various presentative aspects of a material,' was an important variable and one difficult to place under control. Fernald's subjects showed striking individual differences in tendencies toward 'naming,' while studying pictures and other visually-presented materials. Gibson reports that one pattern, *e.g.*, was variously 'thought of' by different ones among his subjects as 'two triangles,' as 'letter W,' and as 'mountains.'

numerals reproduced. An item reproduced correctly both as to form and placement was scored '3' if correct in form but misplaced one position vertically or horizontally it scored '2,' if correct in form but misplaced more than one position a score of '1' was given, and '0' was scored for any item wrong in form or omitted altogether. This is a departure from the form of scoring used by Fernald; it seemed to me to distribute the subjects better and to be better adapted to correlational treatment of the results.

(4) *Arabic Numerals.*—Like the above, except that Arabic numerals only were used.

(5) *Words Alike in Sound but not in Spelling.*—This also formed one of Fernald's 'tests for more accurate analysis,' and like the Roman and Arabic Numerals test this one was intended by Fernald to favor those with strong visual imagery and to handicap those weak or lacking in visual imagery. As in the last two tests the test items are arranged in letter-square form, with ten seconds for study of the visually-presented material. Scoring is the same as above.

(6) *Words Spelled Alike but Pronounced Differently.*—Same as above, except that in this test pairs of words with identical spelling but different pronunciations are used, and the words are read to the subject at the same time that they are presented visually. Time of presentation the same. Intended by Fernald to favor those subjects able to utilize auditory imagery in recall. In the middle position, in each 'form' of this test, is one 'odd' word, unlike in either appearance or sound any of the others.

(7) *Rhyming Ability.*—In the beginning the same list of stimulus words was used which Fernald had used with her subjects. The 'number of rhymes given per unit of time,' did not prove to be diagnostic of imagery differences, but a study of the 'errors' made by the subjects to whom this group of stimulus words was given revealed the fact that most of the subjects making them reported themselves lacking in auditory imagery in the performance of this test. The number of errors made was not large, however, and the test's reliability was low, so an attempt was made to modify the procedure with a view to making the test more difficult and thus distributing the subjects more widely in scores made. A different series of stimulus words was chosen (the original series was 'speak,' 'low,' 'case,' 'home'; the new series was 'lose,' 'cost,' 'loose,' 'boss') which it was believed might necessitate somewhat finer discriminatory capacity, and instead of reading the stimulus-word aloud to the subjects it was now written on the blackboard and left in sight during the minute allowed to find rhymes for it. The instructions laid stress on the fact that the words put down must be good rhymes—that they must *sound* like the stimulus-word. No restrictions were placed on the number of syllables in the words put down, and it was stated that any English words were permissible, including slang, colloquialisms, and proper names, but that words from foreign languages were barred. Various scoring techniques were tried out in connection with this test, and these will be discussed in connection with the presentation of the test results.

(8) *'Childhood' Free Association Test.*—This test was one of the ones used when time did not permit carrying out the regular test schedule. It was given to one section for one half minute, to two other sections for one minute each, and to another for two minutes. The results seemed sufficiently interesting and significant to warrant their being reported. The sections who took this test were simply told that the experimenter would pronounce a stimulus-word, and that they were then to let their minds roam freely and write down every word they thought of until they were told to stop. The word 'childhood' was the stimulus word.

Experimental Results

Under this heading will be presented, first, the results of experimentation considered with reference to the relation between subjects' reports and test scores. Second, test intercorrelations will be presented, for consideration in connection with the question as to the degree of specificity of the abilities measured by the various tests.

I. *Relation Between Test Scores and Subjects' Reports*

1. *Tonal Memory.*—Reports of the subjects as to their manner of performance of this test were grouped into two categories, according as they reported the presence or absence of auditory imagery. The 'auditory' group, in this treatment of the results, thus is made up of all those who reported either 'A' alone, or 'AK' or 'AV' or 'AVK.' The other group comprises all those who reported either 'K' or 'V' or 'VK.' Referring to Table 1, the median score of the 'auditory' group is seen to be 38.67. The median score of the 'non-auditory' group is 29.5. Approximately nine out of ten of the 'auditory' group reach or exceed the median score of the 'non-auditory' group. The difference between the mean scores of the two groups is nine times its probable error, indicating that the difference in favor of the 'auditory' group represents a genuine difference in the ability or abilities which are measured by this test, rather than a difference which might occur by the operation of 'chance' factors alone.

TABLE 1

Comparison of Different Imagery Groups in Scores Received in Tonal Memory and Geometrical Figures Tests

Name of Test	Dominant Imagery	Median	Mean	S.D.	Crit. Ratio	Percent Reaching or Passing Med. of Other Group	Reliability of Test	
							½ vs. ½	Whole Test
Tonal Memory	Auditory ($n=133$).	38.67	37.98	7.68	9.09	87.2	.817	.90
	Non-Aud. ($n=35$).	29.50	29.17	7.54		11.4		
Geometrical Figures	Visual ($n=96$).....	29.04	28.27	3.93	6.65	81.3	.654	.79
	Non-Visual ($n=39$)	25.00	24.48	4.64		20.5		

2. *Test for Memory of Geometrical Figures.*—The subjects were grouped into two classes, as in the test for Tonal Memory, but this time the basis for division was the presence or absence of a 'visual imagery dominant' report. Thus the 'visual' group includes all those who reported either visual imagery alone, or those who reported visual as 'dominant' over whatever else they reported along with it. The 'non-visual dominant' group includes all those subjects who reported either 'K,' 'A,' or 'C' ('C' here stands for reports like 'gave them names,' 'thought of objects,' 'tried to associate them with objects,' and the like), either alone or as dominant over 'V,' and also the subjects who reported 'None' or '?.' Table 1 shows how the two groups compare with each other in their performance of this test. The critical ratio, 6.65, means that the chances are negligible that the difference found between these two groups, in favor of the 'visual dominant' group, could have occurred by chance alone. Eight out of ten of the 'visual dominant' group reach or exceed the median score of the 'non-visual dominant' group.

The conditions under which this test was administered probably weight it in favor of the 'visual dominant' group, and if the conditions were changed in any one or more of several ways the advantage in their favor shown by the above figures would quite probably be wiped out, or even reversed in favor of dominantly 'kinæsthetic' or 'conceptual' performance. That is to say, a five-second presentation of a pair of patterns of the sort used in this test apparently favors those who can 'get' and carry in memory a visual image of the patterns shown, as against those who attempt to utilize the five-second exposure period either for 'kinæsthetic fixation' of the figures or for attempted verbal descriptions of them. But if the patterns shown were more complicated and if scoring were in terms of number of repetitions necessary for perfect reproduction, verbalization and kinæsthesis (except for 'eidetic' individuals, reputedly rare among adults) would perhaps be reported 'dominant' more frequently than 'visual memory' would be, and might result in better scores. The influence of seemingly slight changes in experimental condi-

TABLE 2*

RESULTS FROM FOUR IMMEDIATE MEMORY TESTS

Processes Reported by Subjects	Memory for Roman and Arabic			Memory for Arabic Numerals			Memory for Words Alike in Sound, not in Spelling			Memory for Words Spelt Alike, Pronounced Differently		
	No. Times Reported	Mean Score	Percent of Total Score Obtained from Inaccurate Placements	No. Times Reported	Mean Score	Percent of Total Score Obtained from Inaccurate Placements	No. Times Reported	Mean Score	Percent of Total Score Obtained from Inaccurate Placements	No. Times Reported	Mean Score	Percent of Total Score Obtained from Inaccurate Placements
V.....	70	15.73	9.72	61	15.66	10.78	71	15.73	22.92	31	15.06	17.77
A.....	25	15.32	14.62	35	16.40	6.97	37	14.27	32.39	90	15.33	16.96
VA.....	84	18.39	9.26	61	16.49	6.96	83	15.12	30.68	163	16.81	17.20
VK.....	96	17.84	11.85	88	17.03	7.67	89	15.51	28.91	29	15.24	16.52
AK.....	43	17.63	8.97	55	16.29	8.93	35	14.34	33.01	33	17.58	13.10
VAK....	48	18.88	8.61	50	17.28	9.84	44	16.02	27.23	37	16.76	15.32
No V or A	30	13.73	22.58	30	14.67	15.45	21	11.52	26.86	21	10.62	19.28
	Reliability of Test			Reliability of Test			Reliability of Test			Reliability of Test		
	½ vs. ½ = .399 Whole Test = .57			½ vs. ½ = .314 Whole Test = .48			½ vs. ½ = .564 Whole Test = .72			½ vs. ½ = .584 Whole Test = .74		

* These scores represent the performance of a group of approximately 100 subjects, on four different 'forms' of each test. These results are from the 1928 data; as noted previously the subjects were not then asked to underscore the 'dominant' process. The processes grouped as VAK in the first column include reports of AKV, KVA, etc. Similarly AK includes reports of KA; VK, reports of KV; and VA, reports of AV.

tions upon the reports of subjects who are versatile enough in their command of different imagery modalities to warrant their classification as 'mixed' or 'balanced' with respect to dominant imagery tendencies has been observed by man experimenters.

The objective data, then, in the case of this test as well as the preceding one, shows reports of given imagery processes (which *a priori* analysis would point to as aiding in or essential to the performance of the tests) to be correlated with success, and shows reports of the absence of these processes to be associated with relatively poor test performance in those subjects so reporting.

3. Tables 2 and 3 present data from three of the tests chosen from among the ones called by Fernald 'tests for more accurate analysis of type,' and from one other similar test, 'Memory for Arabic Numerals.'

TABLE 3

RATIO OF THE MEAN SCORE OF THE "NO V OR A" GROUP TO THE AVERAGE OF THE MEAN SCORES OF ALL THE OTHER GROUPS

Arabic		Roman and Arabic		Words Alike in Sound, not in Spelling		Words Spelt Alike, Pronounced Differently	
Total Score	Percent of Total Score Obtained from Inaccurate Placement of Correct Items	Total Score	Percent of Total Score Obtained from Inaccurate Placement of Correct Items	Total Score	Percent of Total Score Obtained from Inaccurate Placement of Correct Items	Total Score	Percent of Total Score Obtained from Inaccurate Placement of Correct Items
.88	1.82	.79	2.15	.76	.92	.65	1.19

The figures listed in Table 2 in the columns headed 'Number Times Reported' show the extent to which subjects of this sort differ among themselves in their reports as to their mode of performance of a given test. The same columns also show the kind and extent of the changes in their reports, with shifts in the kind of material presented or in the manner of its presentation. The first three tests—Memory for Arabic Numerals, Memory for Roman and Arabic Numerals, and Memory for Words Alike in Sound, but not in Spelling—were all presented visually. The fourth—Memory for Words Spelt Alike, but Pronounced Differently—combined visual and auditory modes of presentation.

Some factor, obviously enough, is operating to prevent these subjects all giving the same type of report for any given test, which would tend to be the result if their responses were determined primarily by the mode of presentation rather than by real differences in their manner of attending to the presented material, or in their manner of recall, or both. It is apparent, however, that slight changes in the stimulus-material may for some subjects lead to changes in their reports (which accords with the findings of Segal (**23**), Sybel (**27**), Fernald (**8**), O'Brien (**18**), and others). Note that when the only change in the material presented is that from Roman and Arabic numerals, mixed, to all Arabic numerals, the mode of presentation being visual in each case, the number of 'V' responses drops from 70 to 61 and of 'VK' from 96 to 88, while the 'A' and 'AK' responses increase from 25 to 35 and from 43 to 55 respectively. Observe, too, that the 'Number Times Reported' column shows for all three of the tests involving visual presentation alone a considerable number of reports (28 percent) which fail to include visual imagery, while in the case of the fourth test, with auditory stimulation added to the visual, there are eighty-one reports (20 percent) which make no mention of auditory imagery. It should be noted that the four 'forms' of a given test were never presented one after another. There were always three other tests between any two forms of the same test, so that these shifts in report—some kinds of reports tending to go with one test more than another, as in the case of the increase in 'A' reports in the 'Arabics' test over the 'Roman and Arabic' test—represent genuine tendencies away from stereotypy in the responses given by these subjects.

The same columns also show that the majority of the subjects in the present investigation, like Fernald's more sophisticated group of subjects, would be classed as 'mixed' with respect to their imagery tendencies in test situations like the ones here reported.[6]

[6] In this respect these subjects contrast rather sharply with 'specialized' groups of artists, musicians, scientists, and writers to whom I gave similar tests in an earlier investigation (**5**).

The last test in Table 2 furnishes another point of similarity between Fernald's sophisticated subjects and my own relatively naive group. Fernald states that this test called forth reports of auditory verbal imagery from all her subjects who ever reported such imagery, and the same holds for my group. In fact, all the shifts in reports, and combinations of reports, listed by Fernald as characterizing her group in connection with these tests are paralleled in the accounts given by my unsophisticated subjects; which fact I incline to interpret as lending further support, in addition to that afforded by the objective data, to the validity of the latter's verbal reports.

With reference to objective differences in test scores, between groups giving different reports, perhaps the most significant fact to be noted in Table 2 has to do with the scores received by the 'No V or A' group. In these tests, which were planned to favor the use of either visual or auditory imagery, those who report the absence of both are consistently at the bottom, in mean total score, in all four tests. The consistency indicates that the difference is probably a reliable one, pointing to a real divergence in manner of performance, or in 'ability' of some sort, between this group and the others.

In two of the tests (those calling for the perception and retention of the greater number of discrete impressions) the 'No V or A' group gets a much lower mean 'accuracy' score than do the other groups. One possible explanation of this would be that these individuals, emphasizing vocal-motor processes, take slightly more time for the perception of each discrete stimulus than do those reporting 'visual' and 'auditory' alone or in combination with each other, or with 'K.' This would result in fewer repetitions being possible during the ten second 'learning' period, and accuracy might be lessened in consequence of the fewer repetitions. Another possibility is that a 'perseveration' factor goes with attention to auditory and visual qualities to a greater degree than is the case when attention is dominated by experiences of movement.

The tests involving memory for words give interesting results in another respect. Those reporting neither visual

nor auditory images change their scores for 'accuracy' on these two 'words' tests relatively little with respect to their corresponding scores on the two 'numerals' tests, when compared with the performance of the other groups in this connection. Their mean total scores on the 'words' tests, however, are not only somewhat lower, absolutely, than their own similar scores on the 'numerals' tests, but are also lower, relative to the mean total scores of the other groups. (Table 3 shows the ratio between the mean score of the 'No V or A' group and the average of the mean scores of the other groups combined, with respect to both mean total score and the percent of the total score accruing from inaccurate placements of items correct in form, for all four of the tests here being discussed.) The interpretation already suggested, that those reporting 'No V or A' are handicapped because of the methods which they employ in memorizing and in recall, seems to be reinforced by the results from these 'words' tests. The other groups get approximately the same mean total scores in the 'words' tests as in the 'numerals' tests, but a much larger part of these totals now is contributed by credit given for inaccurate placements of correct items. The 'perseveration' factor, if real, works here for retention of more 'type' or 'key' words rather than for the exact location of many items. With a few words actually 'remembered,' and recorded, the blank spaces in the 'letter-square' can be filled in with words not directly retained as a result of the immediately preceding sensory stimulation, but supplied 'rationally,' since it is known that some of the words are 'like' some of the others. These latter groups, then, reproduced fewer items in the case of the 'numerals' tests than in the 'words' tests but reproduced them more accurately; this is also true of the 'No V or A' group, but to a much less degree.

Fernald believed that tests of the type of the 'Roman and Arabic' and the 'Words Alike in Sound, but not in Spelling' tests would distinctly handicap those stressing auditory imagery to the exclusion of visual, as well as those individuals in whom vocal-motor processes predominate to the exclusion of both visual and auditory imagery. The results recorded

in Table 2 do not show the dominantly 'auditory' individuals, at least, to be under much of a handicap in the performance of these tests, except possibly in the matter of 'accuracy' in the case of the 'words' tests. Here, while the difference in mean total scores is slight between those reporting 'visual' and those reporting 'auditory,' the scores for 'percent of total scores obtained from inaccurate placements' show the 'auditory' groups ('A' and 'AK') much lower in accurate placement of the test items than the 'V' group and somewhat lower than the 'VK' group. In each 'form' of this test there are always three words *different* in sound as well as in spelling, and if any of these three have been retained in memory, through their sound or in any other way, it is possible, as noted above, to 'guess' the other words which sound like them but which are spelled differently, and to insert these at random in the unfilled spaces in the letter-square. But these 'random' insertions cut down the 'accuracy' score, and here visual memory has the advantage, as Fernald intended should be the case. With the 'Roman and Arabic' test, the fact seems to be that the Roman numerals mixed in with the Arabic are quite easily 'verbalized' and remembered in auditory terms by those to whom auditory imagery is to any degree 'natural.' The line '6 IX 4,' for example, instead of being memorized by them as 'six-nine-four' is memorized auditorily as 'six-eye-ex-four' and reproduced accurately in consequence. But in the 'Words Alike in Sound, not in Spelling' test, the short presentation time does not allow words like the following:

'cent	raise	site
sight	raze	scent
rays	cite	sent'

to be spelled out, either auditorily or kinæsthetically, so placement of the separate items, if the words are remembered by their 'sound' or through more or less implicit vocal-motor movements, is necessarily more random than where visual images of the appearance of the words aid in their being correctly localized.

In connection with the test, 'Words Alike in Spelling, not in Pronunciation,' the ratio .65, in the next to the last column

in Table 3, when compared with the corresponding figures for the other three tests (.88, .79, and .76), points to the probability that the pronouncing of the words by the experimenter in this test, along with the visual presentation, was more of a distraction to the 'No V or A' group than it was to the others. If vocal-motor movements predominate in both perception and recall for the members of this group (as they report) the possible inhibiting and distracting effect of simultaneous stimulation of visual, auditory, and proprioceptive sense organs is understandable and checks with the results from previous investigations. (*E.g.*, Fernald (**8**), Dallenbach (**3**), Quantz (**19**), O'Brien (**18**).) The reports from many of our dominantly 'visual' and 'auditory' subjects, in situations like this one, indicate that they exercise the 'attentive selection' among the 'various presentative aspects' of the situation noted by O'Brien. The former attend to the visual impressions and 'shut out' the experimenter's voice. The latter may even shut their eyes and concentrate wholly on the auditory presentation of the stimulus-words, relying on the perseveration of these sound-impressions for success in recall. A possible advantage in favor of those concentrating on the 'visual' method of study, in that they might 'see' the stimulus-words several times while those attending to the auditory stimuli were hearing them only once, might be offset if the auditory sensations perseverated longer, or more clearly. The data for this test in Table 2 reveal no pronounced superiority for one of these groups over the other; except the fact, previously noted, that all the others are better than the 'No V or A' group.

The objective results from these four tests, then, as with the two tests previously reported, may be considered as offering evidence in support of the validity of the subjects' reports; especially since they are in line with objective indications noted by Fernald, using similar tests with more sophisticated subjects.

4. Rhyming Ability Test. This test, with the stimulus-words which were substituted for Fernald's list and with visual presentation instead of auditory, proved highly success-

ful as an objective check upon the reports of the subjects as to the manner in which the words which they put down as 'rhymes' came to them. The clue afforded by the errors made by dominantly 'kinæsthetic' subjects in the first rhyming test given, seems to have been well worth following up. A glance at Table 4 shows how much less discriminating are those subjects who report either kinæsthesis or visual imagery dominant, than those who report that their 'check' upon the words as rhymes is in terms of auditory imagery. Nine out of every ten of the latter make either no errors at all or one only: only one out of the fifty-one subjects in this group makes six errors or more. In the 'K and V' group, there are less

TABLE 4

PERCENT OF EACH IMAGERY GROUP MAKING ERRORS, RHYMING TEST

Number of Errors	Dominant Imagery				
	A (n = 51)	AK, AV, or AVK (None Reported 'Dominant') (n = 18)	K (n = 43)	V (n = 13)	K and V Together (n = 56)
0.....	68.6	38.9	6.9	23.1	10.7
1.....	21.6	11.1	9.3	0.0	7.1
2– 6.....	7.8	38.9	25.6	7.6	21.4
6–10.....	2.0	5.6	25.6	30.8	26.8
10–14.....	0.0	0.0	18.6	38.5	23.2
14–18.....	0.0	5.6	7.0	0.0	5.4
18–	0.0	0.0	7.0	0.0	5.4
	Reliability of Test				
	½ vs. ½ = .86		Whole Test = .92		

than two out of ten, on the average, who make only one error or no errors, while thirty-four out of 56—approximately 61 percent of the whole group—make from six to nineteen errors each. (For list of words scored as errors see appendix, page 657.)

It may of course be argued that the list of stimulus-words chosen was not 'fair'—that it included 'tricky' words. 'Lose,' for example, may be habitually misspelled by some individuals as 'l-o-o-s-e,' and others may be in the habit of

spelling the word 'loose,' 'l-o-s-e,' and the pronunciation may be shifted to go with the wrong words. But as previously pointed out this attempt to confuse was intentional on the experimenter's part and the point of interest is that whatever confusion the experimental situation succeeded in arousing was correlated to such a high degree with reports of a particular sort—namely those of 'V' or 'K' as dominant in the performance of this task. The element of confusion was simply not there for the large majority of those reporting auditory imagery dominant. As a matter of fact, a glance at the list of errors given in the appendix will show that the errors were of many sorts and that all four stimulus words were confusing to some individuals, as it was hoped they would be. Many of these errors seem to indicate that the persons who made them are insensitive to sound-distinctions which are perfectly definite and clear-cut to those individuals reporting auditory imagery dominant in this test.

It is, of course, quite conceivable that there are individuals so sensitive to slight modulations in the proprioceptive stimulation set up by innervation of the vocal-motor muscle groups that distinctions of the sort required in this test could be made by them on the basis of their awareness of very slight variations in kinaesthesis. Three of my 'K' group make no errors and four make only one error. This is sixteen per cent of the entire 'K' group, as contrasted, however, with the ninety per cent of the 'A' group (forty-six individuals out of a total of fifty-one) who find the test such an easy one as such scores indicate it must have been for them.

I find it more difficult to account for three subjects, reporting that they selected the words on a 'visual' basis entirely, who also scored no errors. In their cases it seems to me either that their introspections must be inaccurate, their verbal reports have been entered hurriedly on their papers, and consequently incompletely or inaccurately, or that they *did* choose 'rhymes' purely on a visual basis and by chance happened to get words which 'looked' like the stimulus word and also sounded like it. The last alternative seems improbable, even though the number of words which could be put down in

a minute in response to a given stimulus-word was not very great and the chances for accidental success were consequently better than they would have been if the recall-interval had been several times as long.

A comparison of the results of this test, with these subjects, with the results of the rhyming test used by Stumberg (**25**) with her 'poet' and 'non-poet' groups, is of interest. Her tests differentiated her subjects, in favor of the 'poets,' with respect to the total number of 'right' rhymes given, when her groups were approximately equated for 'intelligence.' No light was shed by her results upon the cause of the difference which she found, and of course the simple fact of unequal amounts of 'practice' might be adequate to furnish the explanation in the case of her subjects. (With the stimulus words used in the present experiment, and the short time-limits, 'total number of "right" rhymes' proved not only a less reliable measure than either 'number of errors' or 'rights minus wrongs' but also a less sensitive index to both 'kind of imagery reported dominant' and 'Alpha' scores.) But the question as to what causes some individuals to 'go in for' poetry, while others with approximately the same amount of 'g' are indifferent to it, is of interest from the purely behavioristic standpoint. Except in the case of free verse, the ability to find 'good' rhymes with ease is probably one of the factors predisposing toward an interest in poetry writing. The verbal reports of the subjects in the present experiment point to the functioning of something which they unite in calling auditory imagery as at least one of the important components of 'rhyming ability.' As in the case of the Seashore Test for Tonal Memory this indication that imagery plays a part in behavior accomplishments of a particular sort raises once more the question which interested Fernald, that of the feasibility of developing the possibly latent imagery potentialities of individuals through 'training.' [7]

[7] From a strictly practical standpoint there are probably many who will hope that we may remain in ignorance on this point, since (such persons will insist) there are too many poets and near-poets turned out even with our present haphazard methods of encouragement of them, and society obviously needs a certain relatively large proportion of non-poets if its more prosaic needs are to continue to be met.

5. 'Childhood' Free Association Test. Results from this test are included here because they bear on the question as to the effect of imagery differences upon overt language responses, and because they seem to be in line with individual differences in association tendencies noted by Stumberg (**26**) and by Downey (**31**). Since only one 'form' of the test was given to the sections which took it no reliability coefficient can be reported. But the results from the several sections which performed the test were consistently in the same direction even though the time limit varied, for different sections, from one half minute to two minutes. The results are tabulated in Table 5.

TABLE 5

FREQUENCY OF NOUNS AND OTHYR PARTS OF SPEECH IN LISTS OF WORDS WRITTEN IN RESPONSE TO THE STIMULUS WORD 'CHILDHOOD'

Time Limit	Mean No. Nouns		Mean No. Other Parts of Speech		Ratios: No. Nouns ÷ No. Other Parts of Speech	
	Reporting V Alone or Dominant	Reporting A, K, VK, AK, or VAK	Reporting V Alone or Dominant	Reporting A, K, VK, AK, or VAK	Reporting V Alone or Dominant	Reporting A, K, VK, AK, or VAK
½ Min.	(n = 22) 9.05	(n = 11) 6.91	1.00	2.18	9.05	3.17
1 Min.	(n = 35) 15.51	(n = 24) 12.21	1.66	6.88	9.34	1.77
2 Min.	(n = 23) 30.09	(n = 6) 19.67	4.39	19.83	6.85	.99

It seems clear that a stimulus-word such as 'childhood' sets off in those individuals who report their associations as exclusively or dominantly 'visual' a tendency to put down words which name the particular scene or individuals 'appearing' to them in their visual images. These subjects consistently put down many fewer 'relational' and 'modifying' and 'action' words than do those who report either something other than visual imagery as dominant, or nothing dominant. The number of other parts of speech written down increases with both these groups as the time limit increases, but the *relative* number of nouns given increases in the case of the

'visual dominant' subjects, decreases with the other subjects, as the response period lengthens.[8]

II. Test Intercorrelations

Tables 6 and 7 show the manner in which the tests already reported upon correlate among themselves and with Alpha. The most striking thing about both tables is that the intercorrelations among the several tests of 'immediate memory' are no higher than they are. In psychological literature, both experimental and theoretical, 'immediate memory' is

TABLE 6

Test Intercorrelations
(1928 Data)

	Arabic	Roman and Arabic	Words Alike in Sound, not in Spelling	Words Pronounced and Shown
Arabic				
Roman and Arabic	.383 ± .06			
Words Alike in Sound, not in Spelling	.106 ± .07	.221 ± .06		
Words Pronounced and Shown	.284 ± .06	.309 ± .06	.466 ± .05	
Army Alpha	.001 ± .07	.229 ± .06	.258 ± .06	.299 ± .06

still frequently referred to as though the term really described some sort of unitary function. The evidence here presented, it seems to me, indicates clearly enough that 'memory' does not operate independently of the kind of material and the manner of its presentation. In the light of the objective support accorded the subjects' verbal reports in the case of the

[8] It might have been anticipated that the 'visual dominant' group might use more adjectives, descriptive of the scenes or persons they were 'picturing' to themselves, than would those whose associations were less concerned with concrete 'particulars.' This did not turn out to be the case. Adjectives were reported in virtually equal proportions by these two groups. There is some slight indication that the 'nouns' of the 'Non-visual dominant' groups tend more toward abstractness than do the nouns put down by the 'visual dominant' groups. That is, there seemed to be a tendency for words like 'justice' and 'kindliness' and 'unhappiness' and 'friendship' to appear more often in the lists of the former than in those of the latter subjects. This tendency might manifest itself more clearly with more subjects and longer time limits, or with subjects of a different sort than these—say college professors, ministers, and lawyers, on the one hand, compared with poets, writers of romantic fiction, painters, and possibly the large and growing tribe of 'mammy-song' writers, on the other.

TABLE 7

TEST INTERCORRELATIONS
(1929 DATA)

	Tonal Memory	Geometrical Figures	Rhymes (R-W)	Words Pronounced and Shown
Tonal Memory				
Geometrical Figures	.201±.06			
Rhymes (R-W)	.132±.06	.195±.06		
Words Pronounced and Shown	.357±.06	.205±.06	.305±.06	
Army Alpha	.291±.06	.335±.06	.622±.04	.508±.05*

* Note: The "words pronounced and shown" test was scored differently in 1929, in that words were scored as correct if they were spelled in accordance with their sound, whether or not this spelling accorded with the words as shown visually. The intention was to weight this test still more in favor of auditory recall. In 1928 only words spelled like the visually-presented stimulus-words were scored as correct. This increased latitude allowed the 1929 subjects may account in part for the higher correlation with Alpha which the 1929 scores in this test yield.

present tests I shall venture to suggest that still other factors which may affect memorial accomplishment are (1) the manner in which a subject 'attends to' or 'perceives' the presented material, and (2) the manner in which he attempts to recall the material.[9]

These tests, however, were specifically devised, or chosen from among tests already in existence, to stimulate different modes of attending and recalling, and to measure their products in case such differences were uncovered. That they succeeded—to the extent indicated by the subjects' reports, by the objective differences in test scores on the parts of subjects reporting different imaginal processes dominant, and by the relatively low intercorrelations—by no means implies that other tests of immediate memory could *not* be found which *would* correlate to a high degree with each other. To find tests of immediate memory which will intercorrelate highly is patently a very simple matter. All that is necessary to

[9] This statement does not, of course, rest on the results from the present experiment alone. The studies of Warden (28), Scott (20), Husband (12), and Cox (2), in the field of human maze learning, the work of Gibson (10), Wulf (30), Messmer (16), Katz (13), and Allport (1), on perception and reproduction of visually perceived forms, and observations, made by Kuhlmann (14), Segal (23), Sybel (17), Meuman (17), Fernald (8), Griffitts (11), O'Brien (18), and others in studies similar in some one or more respects to the present one, all point in the direction indicated.

achieve such a result is to have the materials used in the different tests similar enough in content, and to keep the method of presentation reasonably constant; then inter- and intra-subject differences with respect to modes of perceiving and recalling will scarcely operate to prevent high correlations between the tests.[10]

It is not my intention to enter into any thoroughgoing discussion of Spearman's 'Two-Factor' theory (24), or the various alternative theories to Spearman's, all of which attempt to explain the fact that test intercorrelations tend always to be positive and that they tend to fall into a hierarchical arrangement as regards their magnitudes. The correlational data here presented are inadequate to throw much light on the 'group factor' versus 'g' argument. It is plain enough, however, that if Army Alpha is here taken as a measure of 'g' or 'general intelligence,' the extent to which the latter 'saturates' these tests is for the most part very slight. Whether the abilities called into play in the performance of these tests represent the functioning of specific factors more than they do group factors, or the other way around, is not clear; it *is* quite clear that there is much more involved in their performance than whatever it is that Alpha and similar 'general intelligence' tests measure. (This statement holds even in the case of the 'rhymes' test and the 'words

[10] This is apparent in the reliability coefficients of the tests now under discussion, computed by means of the 'split halves' technique. Except in the case of the Seashore Test of Tonal Memory (in which case the total score from twenty-five 'odd' judgments was correlated with the total score from twenty-five 'even' judgments) the reliability coefficients of all the other memory tests were obtained by correlating scores from two short test 'forms' against the scores from two other similar test 'items' or 'forms.' (With the Geometrical Figures test it was three versus three.) The separate 'forms' of any given test were made up of material which was 'similar' only in some very general way. Still (although, as noted previously, each two successive 'forms' were separated by three other tests) the 'whole test' correlations for the tests treated in this way were as follows: .48 (Arabics), .57 (Roman and Arabics), .72 (Words Alike in Sound, not in Spelling), .74 (Words Pronounced Alike, Spelled Differently, .79 (Geometrical Figures), .92 (Rhymes-Errors), .78 (Rhymes-R-W). These coefficients are not 'high' as reliability coefficients are usually judged, but most of them are *very* high in comparison with the test intercorrelations. For present purposes they may also be considered satisfactory reliability coefficients in view of the shortness of the tests, the homogeneity of the 'population' of subjects, and the fact that the test scores are here used for group comparisons, not for prediction from individual scores.

pronounced and shown' test, which show a higher degree of relationship with Alpha than do any of the others.) The verbal reports of the subjects point to this 'much more' as being in part, at least, different amounts of facility in the use of what these subjects agree to call imagery, of one sort or another.

So far as these tests emphasize retentivity, especially the pure perseveration of sensory stimulation, low correlations with 'g' were to be anticipated. It will be remembered that the early constructors of intelligence tests included tests of memory span on *a priori* grounds, and were somewhat disconcerted to find that these tests not only correlated to a low degree with the other sorts of tests hit upon (which would have been a point in their favor) but that they also correlated low with 'criteria' of general intelligence (which was distinctly *not* in their favor). Now Spearman sums up (**24**) a large amount of work by investigators connected with his laboratory, and announces that their findings all point to a 'retentivity' factor as tending to 'diminish the saturation of a test with "g."' That is to say, whenever a test stresses retentivity, the correlation of that test with measures of 'g' goes down. Obviously most of the tests here being reported stressed retentivity; most of them, intentionally on my part, stressed 'perseveration,' either visual or auditory, or both.[11] The tests which to any extent involve 'eduction of relations,' or of 'correlates,' (the 'rhymes' test, obviously; to a lesser degree probably the two 'words' tests, as already noted at the time the results from these tests were presented and their interpretation considered) are just the ones which according to this theory should be most 'saturated' with 'g.' The intercorrelation tables show this to be the fact. 'Calling up rhymes' may be considered an 'intellectual' task insofar as it involves the *comprehension* of the essential conditions for 'rhyming' and insofar as it also involves having a stock of

[11] For Spearman 'perseveration'—or 'inertia' or 'persistence'—is the 'second division' under the general heading 'retentivity.' These are not hypothetical 'faculties,' as he uses them, but terms descriptive of ways in which test data really behave. The low correlations of my more nearly 'pure retentivity' tests with Alpha are evidently quite in line with Spearman's evidence on this point.

words on hand to select from; it is also 'sensory' or 'imaginal' to the extent that the *check* upon a given word as a 'rhyme' or 'not-rhyme' is dependent upon fineness of discrimination among sounds or movements, sensed or imaged. The former factors tend to push up the correlation with 'g'; the inclusion of demands for fine sensory discriminations in a test tend, like emphasis upon retentivity, to diminish it.

Sex Differences

Griffitts (11) reported small but consistent sex differences in auditory imagery in favor of his women subjects. The test results in the case of the present investigation reveal no consistent, or completely reliable, sex differences. Table 8 presents a comparison of the sexes with respect to their per-

TABLE 8

SEXES COMPARED

(1929 DATA)

Test	*n*	Mean	S.D.	P.E.$_{Mean}$	$M_1 - M_2$	P.E.$_{Diff.}$	$\frac{Diff.}{P.E._{Diff.}}$
Alpha							
Males	48	151.67	20.57	2.00	7.25	2.55	2.84
Females	73	144.42	19.94	1.57			
Visual Designs							
Males	48	28.13	4.37	.43			
Females	75	27.00	4.44	.35	1.13	.55	2.06
Tonal Memory							
Males	48	35.71	9.11	.89			
Females	75	36.81	7.72	.60	−1.10	1.07	1.02
Words Pronounced and Shown							
Males	27	73.30	11.78	1.53			
Females	75	66.65	11.56	.90	6.65	1.77	3.74
Rhymes Right-Wrong							
Males	27	16.11	8.54	1.11			
Females	75	13.59	9.44	.73	2.52	1.33	1.89

formance in the tests used in the 1929 experimental program. (Data from the 1928 tests has not been treated in this way, but the 1929 tests contain more tests in which auditory imagery was reported dominant by a large proportion of the

subjects than was the case with the 1928 series, so any superiority in favor of the women in this respect should show up in this table, if at all.) The column, Diff. ÷ P.E. Diff., shows that none of the sex differences indicated are four times their probable error, the usual statistical criterion of an undoubtedly 'real' difference. The tests showing the highest critical ratios, with the chances in favor of the differences found being 'real' differences, are in favor of the men; but any of the differences reported could be due to chance.

Summary

1. Reports as to manner of performance of various intellectual tasks, by the subjects of this experiment, under the stated conditions of experimental control, are supported to a considerable degree by objective results in the form of differences in test scores correlated with differences in the subjects' reports.

2. Correlations with 'intelligence' as measured by Army Alpha are for the most part positive and low, indicating that there may be something common between the abilities tapped by such 'imagery' tests as those here reported and whatever Alpha measures, but that there is much more which is specific to some or all of the imagery tests, and which is *not* measured by a test of 'general intelligence.'

3. To what extent these abilities which are 'not intelligence' are group abilities, and to what extent they are still more narrowly specific in their behavior, is not clearly demonstrated in the results from the present experiments. The low correlations between Alpha and the immediate memory tests which can be considered most nearly tests of retentivity, or of retentivity plus sensory acuity, and the higher correlations of Alpha with tests like 'rhymes' and the 'words' tests of immediate memory (which seem to involve to a greater or less extent what Spearman calls 'eduction of correlates,' in addition to retentivity and sensory or imaginal discriminative capacity) are in line with results reported by workers in Spearman's laboratory, and may be interpreted as supporting

the latter's hypothesis concerning the independent variability of such factors as 'g,' and 'sensory acuity' and 'retentivity.'

4. The high degree of correspondence noted under (1), above, between test score differences on the one hand, and differences in reports as to the presence or dominance of what most of the subjects seemed to have little difficulty in recognizing as 'imagery' of one modality or another, on the other hand, lends support to the hypothesis that 'imagery' has functional reality. If, that is, the 'image experience,' or the 'image response,' or however you wish to describe it, is shown to have behavior consequences which are objectively demonstrable, then there is no longer any ground for the dismissal of the image as epiphenomenal, unless the anti-imagist can demonstrate a correlation as close, or closer, between the same objective behavior differences and the mechanisms which he observes and which he claims are adequate for purposes of psychological description and explanation.

5. The results of the present investigation with respect to the functional significance of imagery differences are fully in line with observations of other experimenters working in various fields—maze learning, perception, recall, recognition, association, etc.

6. No completely reliable or consistent sex differences are found.

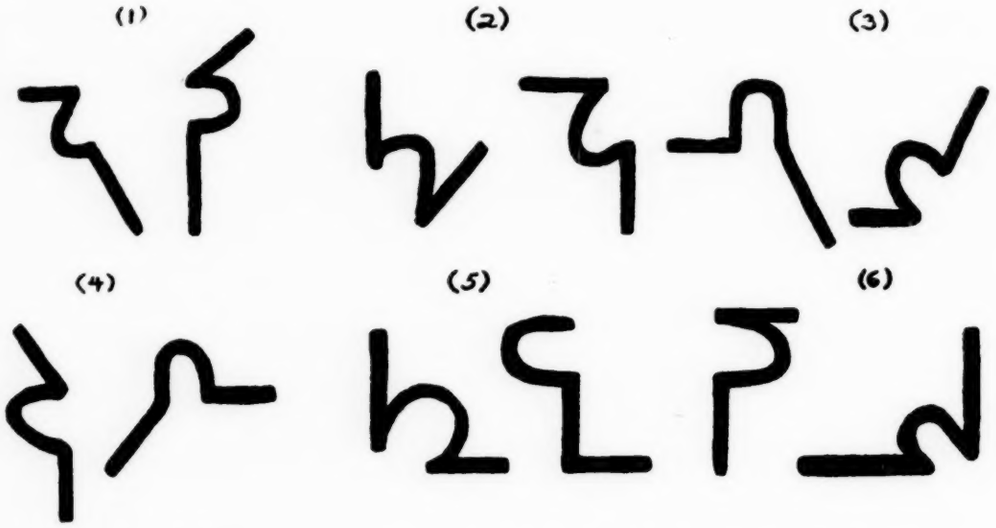
(Manuscript received September 1, 1931)

Appendix

Materials Used in the Tests

Test 1. Test of Tonal Memory.—See Seashore (21) and (22) for a description of the test, directions for administering and scoring, and norms.

Test 2. Memory for Geometrical Figures.—The following pairs of geometrical figures were used:



Each pair of figures was drawn in India ink on a card measuring 6 × 4 inches.

Test 3. Memory for Roman and Arabic Numerals.—The four forms of this test were as follows:

(1)			(2)			(3)			(4)		
6	X	2	3	7	IV	6	V	9	VI	7	XI
V	3	VII	12	V	9	X	3	VIII	4	2	VII
9	8	IV	5	IX	2	2	8	IV	9	X	5

Test 4. Memory for Arabic Numerals.—

(1)			(2)		
11	6	49	9	4	28
3	8	17	73	1	5
5	61	9	2	36	7
(3)			(4)		
3	85	2	4	13	1
6	1	47	7	2	98
8	5	24	5	9	53

Test 5. Memory for Words Alike in Sound, Not in Spelling.—

(1)			(2)		
cent	rays	cite	rain	two	ere
raise	site	sent	too	reign	air
sight	raze	scent	rein	to	heir

(3)			(2)		
toad	sign	mete	right	pair	by
syne	towed	meat	buy	pear	rite
meet	sine	toed	bye	write	pare

Test 6. Memory for Words Spelled Alike, Pronounced Differently.—

(1)			(2)		
			(soo-er)	(wound)	
wĭnd	līves	rōw	sewer	wound	plēad
			(wōōnd)		(sloo)
băss	mall	lĭves	wound	top	slough
			(sluff)		(sow-er)
rŏw	wīnd	bāss	slough	plĕad	sewer

(3)				(4)	
					(teer)
sŏw	pōll	tier	dŏve	bōw	tear
lēad	rust	sōw	bŏw	pool	rēad
		(tie-er)	(tare)		
pŏll	lĕad	tier	tear	rĕad	dōve

Test 7. Test of Rhyming Ability.—The stimulus words were: 'lose,' 'cost,' 'loose,' 'boss.'

Typical 'errors' were as follows:

'lose'	*'cost'*	*'loose'*	*'boss'*
nose	toast	choose	gauze
dose	boast	whose	because
rose	most	booze	cause
moose	ghost	twos	jaws
noose	host	news	saws
dude	post	soothe	morose
lewd	loss	tooth	dose
calaboose	moss	booth	rose
	toss	truth	force
	sought	blouse	norse
	bought	spouse	indorse
	draught	louse	

References

1. Allport, G. W., Change and decay in the visual memory image. *Brit. J. Psychol.*, 1930, **21**, 133–148.
2. Cox, C., Comparative behavior in solving a series of maze problems of varying difficulty. *J. Exper. Psychol.*, 1928, **11**, 202–218.
3. Dallenbach, K. M., Two pronounced cases of verbal imagery. *Amer. J. Psychol.*, 1926, **38**, 667–669.
4. Davis, F. C., Objective psychology and the image. To be published.
5. ——, *The determination of predominant imagery habits in persons of highly specialized training.* 1922. Master's thesis, Univ. of California Library.
6. De Voss, J. C., The specialization of abilities, in *Genetic Studies of Genius*, Vol. I, Stanford Univ. Press, 1925.
7. Downey, J., *The Imaginal Reaction to Poetry.* Univ. Wyoming, Dept. Psychol., Bull. No. 2.
8. Fernald, M. R., The diagnosis of mental imagery. *Psychol. Rev., Monog. Suppl.*, 1912, No. 58.
9. Feuchtwanger, A., Versuche ueber Vorstellungstypen. *Zsch. f. Psychol.*, 1911, **58**, 161–199.
10. Gibson, J. J., The reproduction of visually perceived forms. *J. Exper. Psychol.*, 1929, **12**, 1–39.
11. Griffitts, C. H., Individual differences in imagery. *Psychol. Monog.*, 1927, **37**, No. 3.
12. Husband, R. W., Human learning on a four-section, elevated finger maze. *J. Gen. Psychol.*, 1928, **1**, 15–28.
13. Katz, D., Ueber Individuelle Verschiedenheiten bei der Auffassung von Figuren. *Zsch. f. Psychol.*, 1913, **65**, 161–180.
14. Kuhlmann, F., On the analysis of the auditory memory consciousness. *Amer. J. Psychol.*, 1909, **20**, 194–218.
15. McCarthy, D., A study of the Seashore measures of musical talent. *J. App. Psychol.*, 1930, **14**, 437–455.
16. Messmer, O., Zur Psychologie des Lesens bei Kindern und Erwachsenen. *Arch. f. d. ges. Psychol.*, 1903, **2**, 190–298.
17. Meumann, E., *Vorlesungen zur Einführung in die experimentelle Pädagogik und ihre psychologischen Grundlagen*, II, 1907.
18. O'Brien, F. J., A qualitative investigation of the effects of mode of presentation upon the process of learning. *Amer. J. Psychol.*, 1921, **32**, 249–283.
19. Quantz, J. C., Problems in the psychology of reading. *Psychol. Rev., Monog. Suppl.*, 1897, **2**, No. 5, 1–51.
20. Scott, T. C., The retention and recognition of patterns in maze learning. *J. Exper. Psychol.*, 1930, **13**, 164–207.
21. Seashore, C. E., *The Psychology of Musical Talent.* Boston, 1919.
22. ——, *Manual of Instructions and Interpretations for Measures of Musical Talent.* New York, 1919.
23. Segal, J., Ueber den Reproduktionstypus und das Reproduzieren von Vorstellungen. *Arch. f. d. ges. Psychol.*, 1908, **12**, 124–236.
24. Spearman, C., *Abilities of Man.* London, 1927.
25. Stumberg, D., A study of poetic talent. *J. Exper. Psychol.*, 1928, **11**, 219–234.
26. ——, A comparison of sophisticated and naive subjects by the association method. *Amer. J. Psychol.*, 1925, **36**, 88–95.

27. SYBEL, A. VON, Ueber das Zusammenwirken verschiedener Sinnesgebiete bei Gedachtnisleistungen. *Zsch. f. Psychol.*, 1909, **53**, 257–360.
28. WARDEN, C. J., The relative economy of various modes of attack in the mastery of a stylus maze. *J. Exper. Psychol.*, 1924, **7**, 243–275.
29. WATSON, J. B., *Behavior:* An Introduction to Comparative Psychology. New York, 1914.
30. WULF, F., Beitrage zur Psychologie der Gestalt; Vi. Ueber die Veränderung von Vorstellung (Gedächtnis und Gestalt). *Psychol. Forsch.*, 1922, **1**, 333–373.

THE COMPARATIVE RETENTION VALUES OF A MAZE HABIT, OF NONSENSE SYLLABLES, AND OF RATIONAL LEARNING

BY JOHN A. McGEOCH
University of Missouri

Introduction

The generalization that acts of skill, which are largely nonverbal in character, are better retained than are memorial materials, which are primarily verbal, occurs frequently in the literature of the psychology of learning. It is usually made without qualification. The implication is, therefore, that the skilled acts are intrinsically the more permanent acquisitions. This generalization, if it is valid, has a much greater significance than has been ascribed to it customarily. It is important for theories of mental organization, for theories of the peripheral versus the central character of mental functions, and more immediately, of course, for theories of learning and retention.

In a previous paper,[1] this generalization was questioned on the basis of a review of the literature on the problem and of a direct experimental comparison of the retention values of mazes and of nonsense syllables. We were unable to find in the literature any direct comparison of the retention of skill and of memorial material. The statement that skill is the better retained seems to have been made on the basis of a comparison of results from entirely different experiments which made use of conditions which were far from being comparable. The high degrees of retention found in some of the experiments upon typewriting, ball-tossing, and maze learning, compared with the relatively low degrees of retention yielded by experiments with nonsense syllables and words, seem to have figured largely. In most of such experiments the subjects were few, the experimental conditions differed

[1] J. A. McGeoch and A. W. Melton, The comparative retention values of maze habits and of nonsense syllables, this JOURNAL, 1929, 12, 392–414.

widely and, in particular, the degrees of learning were unequal, having been almost universally greater in the case of skill than in that of memory. There are, moreover, experiments in which memorial materials are forgotten not much more rapidly than are the skills of other experiments. Our review of the literature led to the conclusion that no valid generalization regarding relative retention values can be made from the available experimental work upon the retention of skilled acts and of memorial materials.[2]

An experimental comparison was made of the retention values of mazes of three degrees of difficulty with those of nonsense syllable lists, 8, 12, and 16 syllables in length. Each maze and list was learned to a criterion of one perfect repetition and relearned after one week to a criterion of three perfect repetitions. In terms of an arbitrary criterion, therefore, the degrees of learning of the two materials were the same. The relative retention values were found to be functions of the particular maze and syllable list compared, of the method of measuring retention, and of the relearning criterion employed. The syllables were clearly superior in terms of trials, while in terms of time and errors at least one syllable list was equal to, or better than, at least one maze. There was found, thus, no evidence for the generalization that skill is better retained than is memorial material.

In a recent paper, Freeman and Abernethy[3] have held that this experiment is limited in two respects. (1) The two forms of learning employed differ radically in structure and, as a result, it is not possible to say whether the presence of overt movement, as such, influences retention. (2) Further, "the interval between the learning and the re-learning—one week—is too short to indicate what differences may exist throughout longer intervals." With regard to the first limitation urged, it is, of course, true that mazes and syllable lists differ in structure. In an initial examination of the generali-

[2] For a recent discussion of the problem, see W. S. Hunter, Learning: II. Experimental studies of learning. Chap. 15 in *The Foundations of Experimental Psychology*, ed. by C. Murchison, Worcester, Clark Univ. Press, 1929.

[3] F. N. Freeman and E. M. Abernethy, Comparative retention of typewriting and of substitution with analogous material. *J. Educ. Psychol.*, 1930, 21, 639–647.

zation under discussion, however, it seemed better to employ materials of the kind upon which the generalization was, though in the absence of direct comparability, based. The results of a comparison of such materials under comparable experimental conditions would test the generalization on its own ground and prepare the way for a more analytic study of the specific factors involved. The fact that, under these conditions, no radical differences appear indicates strongly that the greater dominance of motor components does not mean greater retentiveness. It does not, however, constitute direct proof of anything save that, under these conditions, mazes are no better retained than are syllables. The validity of their second point, that a week is too short to be a representative interval, can only be determined by an experimental test. This interval was chosen because most retention curves are falling relatively slowly and uniformly by the time a period of a week has elapsed and there seemed small reason to suppose that what would be true of relative retention values after a week would not be true, in general, after longer periods.

In an experiment designed to overcome these limitations, Freeman and Abernethy had two groups, of 22 subjects each, learn a short paragraph which contained every letter of the alphabet. One group, previously unfamiliar with the keyboard of the typewriter, practiced typewriting the paragraph. The keyes were covered with blank caps and a large diagram was kept on the wall in front of the subject until one errorless copy had been made, after which she practiced without the diagram. The other group translated the paragraph into number symbols. On a small copy of the diagram used by the first group, the letters were numbered according to their position on the keyboard, so that for every movement made in typing there was a corresponding number. After the subject had gone through the paragraph once silently, she recalled what she could, speaking the numbers aloud, and the silent translation with the diagram and the recall aloud alternated until the criterion was reached. At the beginning of each subsequent learning period a recall preceded the silent repetition. Each group practiced in daily periods of 50 minutes

each until a criterion of two successive perfect trials had been reached.

It should be noted that the two materials compared are the same in form but differ in the amount of overt movement to be learned, and that this is an important comparison to make. The conditions of practice of the two groups are not, however, formally the same. The typewriting group practiced with the diagram until one errorless copy had been made and without it until the criterion had been reached. The shift from guidance from the diagram to independent practice introduces a variable the effect of which is unknown. The substitution group had the benefits of a recitation after each independent practice. The report does not state whether the subjects of this group were given immediate information regarding the correctness of their recalls, nor whether those of the typing group saw what they wrote. In the absence of this information it is difficult to estimate the extent to which the different methods of practice may have affected the results. The typewriting group required roughly 46 minutes, or one practice period, more to reach the criterion than did the substitution group. This introduces both more practice, in terms of time, and a distribution over a longer time. Freeman and Abernethy, however, consider this unimportant.

After two weeks each group relearned the material with the same procedure as in learning. Eight weeks after this relearning it was again relearned as before. The saving scores, in terms of errors, repetitions and time are very nearly the same for the two materials at the first relearning. This is cited as being in agreement with the results of McGeoch and Melton. At the second relearning, ten weeks after the original learning and eight weeks after the first relearning, different results appear. The saving scores of the typewriting group show a gain over the corresponding scores at the two-weeks relearning, while those of the substitution group show a loss. Freeman and Abernethy write: 'The difference is unmistakable. It seems to confirm without question the traditional belief that, for longer periods of time, learning which involves overt movements is more permanently retained than is learning which

does not involve overt movements. The difference is so great that deterioration of the overt habit seems actually to be halted after two weeks.' This is cited as the most important result of the experiment and, had it appeared under unequivocal conditions, it would have been of the first importance.

Two facts, however, render very doubtful the interpretation of the saving scores after ten weeks. (1) Each group relearned after two weeks and again after ten weeks. The second relearning has, thus, had antecedent to it both the original learning and the relearning after two weeks. (2) At the first relearning, the typewriting group required roughly 86 percent more time for relearning than did the substitution group, although the saving scores were roughly the same.

The superior retention of the typewriting after ten weeks may mean that learning which involves overt movements is better retained than learning which does not involve such movements. There are, however, two other interpretations which must be considered and which are very possible ones. It may be that learning involving overt movements is the more effectively reinstated and refreshed by a distributed practice (relearning after two weeks). It may be also that the greater amount of time taken to relearn the typewriting was a factor in favor of its greater retention at the second relearning. Against this it may be argued that the greater time taken to learn the typewriting gave it no advantage at the first relearning and hence that the greater time taken to relearn it the first time should give it no advantage at the second relearning. This is possible, but it is also possible that the effect of greater time spent is cumulative or is more potent at relearning than at learning. It may be noted that the relative disparity between the two materials, in terms of the means, is virtually the same at the first relearning as at the original learning. Any one of the interpretations suggested above would be important, if established. Under the conditions it seems to the writer impossible to guess, with any factual support, which is most likely. The only appeal is to experiment.

It is the implication of the interpretation of Freeman and Abernethy that, had we carried our experiment with the

mazes and nonsense syllables over a sufficiently long period, we would have found the mazes to be the better retained. If valid, this would be highly significant and would necessitate a radical reinterpretation not only of our results, but of curves of retention for motor skills. Two experiments have, therefore, been performed to test this and to study, as well, the influence of an interpolated relearning.

Experiment I

Method.—In this experiment two of the six materials, the easy maze and the 8-syllable list, used in the McGeoch and Melton experiment have been employed. Group I, composed of 22 college students who were entirely unfamiliar with maze learning or with nonsense syllable learning by any method, learned each material to a criterion of one perfect trial and relearned it after a period of eight weeks [4] to a criterion of three perfect trials in succession. Group II, composed of 18 subjects of the same kind, learned each material to a criterion of one perfect trial, relearned it after one week to the same criterion, and again relearned it after an interval of eight weeks from the original learning, this time to a criterion of three successive perfect trials. In each group each subject learned both materials. Half of the group learned the maze on Day 1 and the syllables on Day 2. Half reversed this order. The order of relearning was, of course, dictated by the order of learning. The maze was screened from the subject and the usual instructions were employed. The syllables were exposed at a 2-second rate on an electrically driven memory drum and were learned by the anticipation method. In every way the conditions of the previous experiment were duplicated as closely as possible. The same precautions against rehearsal were used, and the results have been treated in the same say.

Results.—The mean learning, relearning and saving scores for Group I, composed of 22 subjects who relearned only once, after 8 weeks, are given in Table I. The time scores are in seconds.

The saving scores are uniformly higher for the nonsense syllables than for the maze habit. The superiority of the syllables is greatest in terms of trials. The difference between the two materials is much less when saving is computed in terms of time and errors, but it is still considerable, and is greater in terms of errors than of time. In learning syllable lists by the anticipation method, the time score is the trial

[4] A duplication of the 10-weeks interval used by Freeman and Abernethy was at first intended, but it turned out to be impossible to get all subjects through the original learning in time to allow more than 8 weeks to elapse without encroaching on the final examination period, with the variable conditions which that would imply. The difference between 8 and 10 weeks can scarcely affect the general implications of the results.

TABLE I

GROUP I: LEARNING, RELEARNING AND SAVING SCORES

Maze Scores

	Learning		Relearning—8 weeks		Saving
	Mean	σ_M	Mean	σ_M	
Trials.....	17.18	2.33	30.50	5.33	−77.53
Time.....	572.09	86.99	437.45	78.03	23.53
Errors.....	130.81	28.99	90.68	21.64	30.67
Nonsense Syllable Scores					
Trials.....	24.18	2.50	15.36	1.67	36.47
Errors.....	119.04	11.68	53.72	5.69	54.87

score multiplied by the constant exposure time for the total list. The saving score is, thus, the same for time as for trials.

The results uniformly corroborate those of the prior experiment. Nonsense syllables are retained somewhat better than is the maze habit after eight weeks as well as after one week. In the present experiment only two of the former materials have been used, but there is no reason to believe that these two are not fair samples.

A repetition of this experiment under conditions analogous to those employed by Freeman and Abernethy will assist in an interpretation of their results and will show the influence of an intervening relearning. Group II learned the same materials as did Group I, but relearned them after one week and again after a total of eight weeks. Any major differences between the tendencies of the saving scores of the two groups may be ascribed to the interpolated relearning of the second group. The first relearning is introduced after one week to conform with the interval of the prior experiment. The results yielded by this group are given in Table II.

It will be noted that the means of the learning records of Group II are somewhat larger than are those of Group I. This difference occurs with both materials, and evidence will be presented later that the relative retention values have not been altered thereby. The larger scores of Group II are produced by a very few slow learners who happened to fall

TABLE II

GROUP II: LEARNING AND RELEARNING

Maze Scores

	Learning		Relearning—1 Week		Relearning—8 Weeks	
	Mean	σ_M	Mean	σ_M	Mean	σ_M
Trials.....	24.44	4.22	10.05	2.72	19.50	3.82
Time.....	661.38	107.07	161.94	35.51	216.72	40.33
Errors.....	153.27	23.70	36.72	11.07	35.22	8.55
Nonsense Syllable Scores						
Trials.....	27.66	3.86	5.55	.76	16.33	2.09
Errors.....	140.72	20.47	23.50	3.19	53.33	7.61

in this group and whose records increase the means and the variability.

The saving scores computed from the data of Table II are shown below. The saving at the relearning after 8 weeks has been computed in terms of the sum of the records for learning and for relearning after one week. Since both are required to represent the subject's antecedent practice with the materials, both have been included in the denominator in finding the saving. The criterial trials of the relearning after eight weeks are not, of course, included in the numerator. The relations between the two materials are in no essential way altered, however, by computation of saving scores with only the learning records in the denominator.

TABLE III

GROUP II: SAVING SCORES

	Relearning—1 Week			Relearning—8 Weeks		
	Trials	Time	Errors	Trials	Time	Errors
Maze.....	58.87	75.51	76.04	45.07	73.70	81.56
N.S.....	79.93	79.93	83.30	52.27	52.27	67.52

The saving scores after the interval of one week show essentially the same relationships as do those of the preceding experiment. The superiority of the syllables is clear when the measurement is in terms of trials. In terms of time and of

errors the syllables are still superior, but by amounts too small to warrant any conclusion save that the two materials are retained almost equally well when retention is measured in these terms.

When, however, we turn to the saving scores after 8 weeks, a different picture presents itself. In terms of trials the syllable list is still superior, but to a much smaller degree. In terms of time and errors the maze is superior by fairly substantial amounts. These results must be interpreted in comparison with those of Group I which relearned only after 8 weeks and which yielded a clearly greater permanence for the syllables than for the mazes. The only experimental difference between the two groups lies in the relearning after one week, which was absent in Group I and present in Group II. It may, then, be concluded that an intervening relearning one week after the attainment of the arbitrary criterion of one perfect trial, strengthens the maze habit more effectively than it does the verbal habits of the nonsense syllable list. This may be a result either of a differential influence of the distributed practice involved in the interpolated relearning, or of the relatively greater relearning effort required by the maze, or of a combination of the two factors.

At this point it is necessary to inquire whether the results have been affected by unequal amounts of effort in learning. A comparison of Tables I and II will show, not only that the means of the learning records of Group II are larger than those of Group I, but also that the difference is greater with the maze than with the nonsense syllables. There is a considerable amount of evidence in the literature to indicate that retention is a function of the amount of learning effort expended. It might be, therefore, (1) that the relative increase in the retention of the maze habit at the relearning after 8 weeks, in the case of Group II, is a function of the relatively greater time, trials and errors required by that group to learn, and (2) that the tendency of the maze habit to be retained better than the nonsense syllables, when there has been an intervening relearning, may be a function of these higher learning scores rather than of this first relearning.

These possibilities have been investigated by selecting groups which are paired in some measure of learning. Which learning score is the better measure of effort is not certainly established, but time seems a representative one in the case of these problems. Accordingly, to test the possibilities mentioned, 15 subjects, paired as closely as possible in time required to learn the maze, have been chosen from each group. The mean scores differ from each other by only 18 seconds, and the trial and error scores for the same subjects differ much less than do those for the original groups. The same pairing has been done for nonsense syllables in terms of trials (which are the equivalent of time scores when the exposure time per trial is a constant). The means of the trials for the two groups of 15 subjects are identical and the means of the errors differ by only three points. The saving scores for these paired groups are given in Table IV.

TABLE IV

SAVING SCORES OF GROUPS EQUATED IN TERMS OF TIME REQUIRED FOR LEARNING

Group I

	Relearning—1 Week			Relearning—8 Weeks		
	Trials	Time	Errors	Trials	Time	Errors
Maze.....				−86.7	22.0	33.7
N.S.....				45.4	45.4	59.8
Group II						
Maze.....	48.4	68.0	69.9	25.6	63.9	75.6
N.S.....	76.3	76.3	79.0	50.9	50.9	64.5

The same relationships are found with these as with the original groups. When there has been no relearning after one week, the syllables are uniformly the better retained. They are likewise superior when relearning occurs after one week. At the second relearning, after a total of 8 weeks, however, the syllables are superior only in terms of trials, while in terms of time and errors superiority passes to the mazes. Since there has been an equation of groups in terms of time spent

in learning, and a close approximation to equation in terms of the other scores, and since the syllables are uniformly the better retained after 8 weeks with no intervening relearning, it is necessary to conclude that the shift toward a better retention of the maze in the case of Group II must be ascribed to the intervening relearning. This has affected the maze habit much more strikingly than it has the verbal habit.

The method of pairing used has equated the time scores on the maze in Groups I and II and, likewise, the trial (time) scores on the nonsense syllables. That is, it has equated Groups I and II on each problem but has not equated problems. It is hardly fruitful to discuss in this context the problem of equating two such widely differing problems, and the conclusions drawn do not require it. As one additional check on the results, however, it will be useful to attempt an equation of both groups and problems. This has been done in terms of trials. Ten subjects have been found whose trial records on the maze in Group I can be approximately duplicated by subjects on the maze in Group II and in the nonsense syllable learning of each group. There are, thus, available four groups of 10 subjects each whose mean trial scores are virtually identical. The saving scores of these subjects, given in Table V, show the same relationships which were found when the total groups were considered. The agreement of the saving scores obtained from equated sub-groups with those from the total groups removes the possible interpretation of the results as due to errors of sampling and their attendant factors.

TABLE V

SAVING SCORES IN TERMS OF TRIALS FOR EQUATED GROUPS OF TEN SUBJECTS

	Relearning—1 Week		Relearning—8 Weeks	
	Maze	N.S.	Maze	N.S.
Group I.....	65.3	71.9	−71.5	35.9
Group II.....			36.5	53.0

While results from mazes and syllable lists are not known to be applicable without modification to typewriting and substitution, as used by Freeman and Abernethy in their experi-

ment, or to skill and memory in general, it is a plausible hypothesis that in their experiment the superior retention of the typewriting after ten weeks, although it had been no better than the substitution after two weeks, is a function of the intervening relearning. The tentative generalization implied here is that an interpolated relearning strengthens acquisitions involving overt movement more than it does those of the memorial order. Whether it operates directly or through the factor of unequal practice at relearning cannot now be determined. Granting its truth, it might be argued that this supports the generalization that skilled acts have greater intrinsic permanence. As usually stated, the latter generalization applies only to the fate of acquisitions over an interval during which no practice is supposed to have occurred, and in this form the hypothesis offered neither supports nor affects it in any way. If this hypothesis is correct, it may account for the high retention values found after long intervals in the case of many acts of skill. Most of these skills were learned over a long period of distributed practice, and in several cases were given one or more relearnings prior to the final measurement after a relatively long interval. If it is correct, it offers a key to the understanding of several of the phenomena of retention. As it stands, it appears to be true in the case of a maze versus a syllable list. It is tentatively assumed that the result may be generalized to cover the classification skill versus memory.

The results of this experiment are directly at variance with the usual generalization that skill is retained better than is memorial material. After one week and after eight weeks, when no intervening relearning occurs, the syllable list is better retained than is the maze habit. Its superiority is considerable in terms of trials, slight in terms of time and errors. When, however, each material is given an interpolated relearning after one week, the superiority of the syllabels is, at the second relearning after eight weeks, diminished in terms of trials and passes to the mazes in terms of time and errors.

It will at once be evident to anyone who compares the means

and saving scores of this experiment with those of the previous study that there are wide differences. The relative retention values of the materials compared are the same in tendency, but the magnitudes of the scores from which savings are computed vary, as do the magnitudes of the saving scores. This might seem at first consideration to render doubtful the comparability of the conditions of the two experiments. There is, however, one factor which accounts for the difference and which renders it unimportant in so far as the present conclusions go. In this experiment, where only one sample of each material has been used, there is little opportunity for transfer effects to occur. In the former experiment, three samples of each material were used and were learned in a counterbalanced practice order such that one sample of each was learned first of that material by eight subjects, second by eight others, and third by eight others, the whole scheme working out to give all samples equal average position. This arrangement renders inevitable the possibility of both positive and negative transfer from sample to sample and is almost certain to produce results which are different in magnitude from those which would appear if each sample were learned in complete isolation from any influence by others.[5] One would not, however, expect the tendencies of an experiment in which all samples are included and rotated in a counterbalancing order to be different from those of an experiment in which each sample is learned independently by a separate group of subjects. The fact that the relative retention values of this and the former experiment agree so completely constitutes effective proof that the two methods can yield similar tendencies.

[5] Evidence of the influence of transfer effects within the counterbalanced order may be cited from the data of the 1929 experiment. In this experiment the eight subjects who learned the easy maze first, with no previous maze experience, required an average of 16.62 trials, 676.37 seconds, and 193.62 errors. The eight who learned this maze in the last position in the practice order, with prior experience of learning and relearning two other mazes, required 5.75 trials, 127.37 seconds, and 34.75 errors. Obviously the practice effects are very large. The saving scores are much larger at the first position in the practice order than at the last position. This is in agreement with the influence of transfer upon retention in the case of rational learning. Cf. J. A. McGeoch and V. J. Oberschelp, The influence of length of problem and of transfer upon rational learning and its retention, *J. Gener. Psychol.*, 1930, 4, 154–168.

Experiment II

In this experiment a comparison is made between the retention values of a Warden 10-alley stylus maze,[6] and a 10-letter Peterson Rational Learning Problem,[7] after intervals of one week and three weeks. Each material was learned at a single sitting to a criterion of one perfect trial. Half of the subjects learned one material on Day 1 and the other material on Day 2, and the other half reversed this order. Group I, consisting of 20 college students, relearned to a criterion of three successive perfect trials after an interval of three weeks. Group II, consisting of 20 college students, relearned after one week and again after three weeks, each time to a criterion of three successive perfect trials. All subjects were, at the beginning of the experiment, entirely unacquainted with the problems used, and all were given the usual instructions. The same precautions against rehearsal were observed as in Experiment I. The scoring of the rational learning problem in terms of trials, time, logical, perseverative and unclassified errors is well known. For the present purpose the three kinds of errors have been included in a single mean. The instructions used by Peterson were employed. Errors in maze learning were counted as in Experiment I and in the 1929 experiment.

The data yielded by Group I are given in Table VI. In this and in the following table the mean learning scores include the criterial learning trial; the mean relearning scores do not include the criterial trials. In computing saving scores, however, the criterial trials and time of all preceding learning (and relearning, if any) have been included in the denominator

TABLE VI

Group I: Learning, Relearning and Saving Scores

Maze Scores

	Learning		Relearning—3 Weeks		Saving
	Mean	σ_M	Mean	σ_M	
Trials	20.65	2.12	15.75	1.20	23.72
Time	1062.95	276.57	604.95	59.81	43.08
Errors	83.55	11.71	57.85	6.12	30.76

Rational Learning Scores

Trials	5.50	.58	2.85	.31	48.18
Time	464.95	46.30	232.95	26.77	49.87
Errors	72.60	10.29	29.70	4.57	59.09

[6] C. J. Warden, Primacy and recency as factors in cul-de-sac elimination in a stylus maze, this Journal, 1924, 7, 98–116.

[7] J. Peterson, Experiments in rational learning, *Psychol. Rev.*, 1918, 25, 443–467. The problem used in the present experiment was:

A	B	C	D	E	F	G	H	I	J
4	7	1	9	2	10	6	5	8	3

but the criterial scores of the particular relearning at which the saving is being computed have not been included in the numerator. It is obviously necessary to take into account all antecedent practice, whether criterial or not, but it is not necessary to include the criterial trials of the particular relearning because the material is assumed to have been relearned by the time the criterion can be met and the criterial trials may be thought of as tests of the relearning.

In terms of trials and of errors, the rational learning problem is much better retained than is the maze. In terms of time, the difference is small, with the rational learning problem still superior.

Group II in this experiment follows the pattern of the corresponding group in Experiment I, but with its second relearning after an interval of three weeks. The 20 subjects relearned after one week and again after three weeks. The results appear in the following tables.

TABLE VII

GROUP II: LEARNING AND RELEARNING

Maze Scores

	Learning		Relearning—1 Week		Relearning—3 Weeks	
	Mean	σ_M	Mean	σ_M	Mean	σ_M
Trials.....	28.85	3.81	14.20	1.84	8.65	1.75
Time.....	1630.70	164.52	521.65	83.11	264.70	50.87
Errors.....	134.55	15.06	39.00	6.35	19.25	3.40
Rational Learning Scores						
Trials.....	6.60	.55	2.40	.37	1.30	.22
Time.....	640.00	57.68	160.55	34.63	74.00	13.79
Errors.....	95.15	16.21	18.60	5.61	8.45	2.49

TABLE VIII

GROUP II: SAVING SCORES

	Relearning—1 Week			Relearning—3 Weeks		
	Trials	Time	Errors	Trials	Time	Errors
Maze.....	50.77	68.01	71.01	81.21	88.12	88.90
Rational L...	63.63	74.91	80.45	89.16	91.57	92.57

The saving scores after the one-week interval, though larger than those after the three-weeks interval of Group I, show similar relationships. By every measure the rational learning problem is retained better than is the maze. Several of the differences are small, but their consistency warrants the conclusion that after one week and after three weeks the rational learning problem is better retained than is the maze habit.

The saving scores of Group II after the three-weeks interval are uniformly higher than after the interval of one week. They are much higher than the corresponding scores of Group I which had no relearning after a week. Since the variable condition is the interpolated relearning, we must conclude that the height of the saving scores of Group II after the longer interval is a function of the relearning after one week.[8]

Important, also, for the present problem is the fact that, after the interval of three weeks, the saving scores of the two materials approach each other more closely than after the one-week interval. This means, as inspection of Table VIII will show, that from the standpoint of retention after three weeks the maze has profited more from the relearning after one week than has the rational learning problem. This difference in the extent to which the two materials profit from the interpolated relearning is not as great as that which appeared in Experiment I, but it is exactly the same in tendency. In the present experiment the interval is shorter, the rational learning problem is substituted for the syllables, and the magnitudes of the scores compared are different. These factors are sufficient to account for the smaller differences found.

The rational learning problem is one in which ideational factors figure largely and in which no overt movements other than the habituated acts of speaking the numbers are directly

[8] It is possible that the greater learning effort required by this group, as represented by their higher learning records, might have contributed to its better retention after three weeks. As a check upon this the means for the original learning of Group II were made to approach those of Group I by the elimination of the extreme cases. The saving scores were then computed from the new means. The magnitudes of the scores are slightly changed but the relationships remain unaltered.

involved. The greater permanence of this problem compared to that of the maze habit, after both one and three weeks, when no relearning intervenes, constitutes further evidence against the current generalization that skilled acts, involving overt movements, are better retained than are verbal and memorial materials. The fact, moreover, that the two materials are considerably closer together in saving values after three weeks, when a prior relearning has occurred than when it has not, implies that a relearning one week after the attainment of the arbitrary criterion of one perfect trial strengthens the maze habit more effectively than it does the rational learning. This directly parallels the corresponding conclusion of Experiment I and supports the interpretation offered there of the results of Freeman and Abernethy.

Summary

The retention values of the easy maze and the 8-syllable list of the McGeoch and Melton (1929) experiment have been compared after an interval of eight weeks, both when no intervening relearning has occurred (Group I) and when there has been a first relearning after one week (Group II).

1. When no intervening relearning occurs, the nonsense syllables are uniformly somewhat better retained than is the maze habit. The superiority of the syllables is greatest in terms of trials, but is still considerable in terms of both time and errors. The failure of these results to substantiate the current statement that skilled acts are better retained than are memorial materials corroborates the outcome of the 1929 experiment and fails to fulfil the implication of the results of Freeman and Abernethy.

2. In Group II, the saving scores at the first relearning, after one week, show essentially the same relationships as do those of the 1929 experiment and constitute an independent corroboration of the results of that experiment. The saving scores after eight weeks, when a first relearning after one week has preceded, present a different set of relationships. In terms of trials the syllable list is still superior but to a much smaller degree. In terms of time and of errors the maze is

superior by substantial amounts. It is concluded that this difference between the relative retention values of Groups I and II is a function of the relearning after one week in the case of Group II, and that this relearning has strengthened the maze habit more effectively than it has the nonsense syllable list. It is highly probable, in the light of this result, that the superior retention, after ten weeks, of the typewriting in the Freeman and Abernethy experiment, although it had not been superior after two weeks, is a function of the first relearning, after two weeks, and not of the inherently greater permanence of motor activities after longer intervals.

A similar comparison has been made of the retention values of the Warden 10-alley maze and a 10-letter Peterson Rational Learning Problem after an interval of three weeks, both when no intervening relearning has occurred (Group I) and when there has been a first relearning after one week (Group II).

1. When no intervening relearning occurs, the rational learning problem is better retained than is the stylus maze.

2. In Group II, the saving scores at the first relearning, after one week, show a corresponding and consistent superiority for the rational learning problem. At the second relearning, after three weeks, the rational learning problem is still the better retained of the two, but the differences are considerably smaller than in Group I where no intervening relearning occurred. This supports the interpretation offered in the case of Group II of Experiment I.

The results of these two experiments agree in finding that materials which are verbal and ideational in character are, when learned to a uniform criterion, uniformly retained, over both relatively short and relatively long intervals, better than are maze habits. They agree, also, in showing that a relearning after one week strengthens the maze habit more effectively than it does the verbal and ideational acquisitions, although the factors producing this differential reinforcement cannot be determined under the present experimental conditions. They agree, finally, in contradicting, within the limits of the materials and methods employed, the current generalization that skilled acts are better retained than are memorial materials.

Although these experiments have found the memorial materials used to be retained better than the acts of skill with which they have been compared, it is not assumed that the memorial materials possess any greater intrinsic permanence. It is more probable that their superior retention values are a function of experimental conditions not isolated in these experiments.

(Manuscript received August 15, 1931)

MENTAL SET IN RELATION TO RETROACTIVE INHIBITION [1]

BY OLIVE P. LESTER
The University of Buffalo

SECTION I. INTRODUCTION

The fact that the recall of any learned material is interfered with or decreased by the interpolation of certain forms of activity between the original learning and recall has been clearly demonstrated in numerous investigations. Psychologists have denoted this occurrence by the term retroactive inhibition. The investigation under consideration assumes the existence of this retroactive phenomenon and then raises the question as to the possible relation between inhibition and learning set or attitude. More specifically stated, what is the influence or effect of the following attitudes on the susceptibility of learned material to retroactive inhibition?

1. The expectation of recall.
2. The expectation of interpolated material before recall.
3. Information as to the possible effects of interpolated material.
4. The assumption of an active attempt to avoid the detrimental influence of the interpolated material.

Space does not permit an historical review of the experimental literature on retroaction, and attitudes at the time of learning. To the writer's knowledge, no work on the relation of mental set to retroactive inhibition has been reported, though much experimental data on set in relation to learning, and recall is at hand.

SECTION II. PROCEDURE

A. MATERIAL

The memory material employed throughout the investigation, both for the original and for the interpolated learning, consisted exclusively of standard lists of twelve three-letter nonsense syllables.

---

[1] From the Psychological Laboratory of The University of Chicago. The writer wishes to express her appreciation to Professor A. G. Bills for the helpful advice and assistance given by him throughout this experimentation.

B. Apparatus

An electrically driven memory drum, of the type devised by Clark Hull, constituted the apparatus. The lists of nonsense material were placed on the drum, and the timing device set so as to allow a constant exposure of three seconds for each nonsense syllable. Thus, a complete presentation of the twelve syllables necessitated thirty-six seconds. The drum continued to rotate until each subject was able to give one perfect recitation of the list, *i.e.* to anticipate by spelling aloud, in each case, the syllable following the one then being presented to him on the drum. The list of nonsense material was then considered learned. The automatic control of the presentation time left the experimenter free to record the learning data of the subjects.

C. Subjects

The subjects were undergraduate students at the University of Buffalo during the college year 1929–30. The data in the main were obtained from freshman men and women, although the results on a number of second year students were also included. In those instances in which sophomores were used as subjects, only those who were not carrying the regular elementary course in psychology were included; or, if such students were employed, the experimenter made certain, in each case, that class discussions had not, up to that time, involved any material on retroactive inhibition, set at time of learning, and the like. This precaution insured that very little, if any, knowledge existed on the part of the subjects concerning the problem under investigation.

D. Methods

The experiment was performed to answer the question whether directions at the time of learning (set) would have any effect upon the susceptibility of learned material to retroactive inhibition. Such a problem called for the usual learning and retention set-up, *i.e.* the learning of the original material; a time interval between original learning and recall, during which the interpolated material was introduced; the subsequent recall of the originally learned material.

In this experiment, the time interval between learning and recall was twenty-four hours; the interpolated learning material was introduced immediately prior to recall. The retention of the original material was measured by the anticipation and relearning methods, *i.e.* the number of nonsense syllables correctly anticipated during the first trial of the relearning constituted one measure of retention, and the total number of trials necessary to relearn the list to the point of one perfect anticipation of each of the twelve succeeding nonsense syllables was taken as the second measure of retention.

Simple, uniform directions for the recall technique were given to all subjects just prior to recall.

The original learning, the time interval, the interpolated learning, and the recall procedure remained the same throughout the experiment. Different directions read by the subject at the time of learning introduced the factor of varying set or attitude —the one variable in our experimental investigation. These different directions, as presented to the subjects, determined the specific conditions of the experiment which were as follows:

Condition I

Learn O[2]	No I[3]	Recall
No expected recall		

[2] Abbreviation for original learning.

[3] Abbreviation for interpolated learning.

Condition II		
Learn O	No I	Recall
Expected recall		
Condition III		
Learn O	I introduced	Recall
No expected recall		
No expected interpolated task		
Condition IV		
Learn O	I introduced	Recall
Expected recall		
No expected interpolated task		
Condition V		
Learn O	I introduced	Recall
Expected recall		
Expected interpolated task		
Condition VI		
Learn O	I introduced	Recall
Expected recall		
Expected interpolated task; S. told something of its possible effects		
Condition VII		
Learn O	I introduced	Recall
Expected recall		
Expected interpolated task; S. told something of possible effects, and urged to resist effects		

The printed directions, read by the subjects at the time of learning, with reference to the expectation of recall and interpolated learning, were as follows:

Condition I

No printed directions were given at the time of learning the original material, but the subject was instructed to return the following day for the purpose of learning another list. This second learning was actually carried out, but only after the recall of the original list.

Condition II

"You are going to be required to recall, tomorrow at this time, the list of nonsense syllables you are now learning. Do not practice on the list after leaving here."

Condition III

No printed directions were given at the time of learning the original list, but the subject was instructed to return the following day in order to learn another list. The learning of this list on the succeeding day, since it constituted, in reality, the interpolated task, was actually carried out, after which the recall of the original material took place.

Condition IV

"You are going to be required to recall, tomorrow at this time, the list of nonsense syllables you are now learning. Do not practice on the list after leaving here."

These directions were identical with those of Condition II.

Condition V

"You are going to be required to recall, tomorrow at this time, the list of nonsense syllables you are now learning, but, before recalling this list tomorrow, you will learn another list of nonsense syllables similar to the list you are learning today. Do not practice on the list you learn today after leaving here."

Condition VI

"You are going to be required to recall, tomorrow at this time, the list of nonsense syllables you are now learning, but, before recalling this list tomorrow, you will learn another list of nonsense syllables similar to the list you are learning today. The effect of learning a second list is usually detrimental to the remembering of the first list—that is, it is likely to cause confusion. You tend to mix up syllables learned in the second list with those in the first list. Do not practice on the list you learn today after leaving here."

Condition VII

"You are going to be required to recall, tomorrow at this time, the list of nonsense syllables you are now learning, but, before recalling this list tomorrow, you will learn another list of nonsense syllables similar to the list you are learning today. The effect of learning a second list is usually detrimental to the remembering of the first list—that is, it is likely to cause confusion. To avoid mixing up the syllables of the second list with those of the first list, take special care to keep them separate. Learn the first list in such a way that you will not confuse it with the second list. Do not practice on the list you learn today after leaving here."

In order to reduce to a minimum the variability necessarily accompanying oral directions given by an experimenter, each subject was requested to read the directions, and then to state his understanding of them. In the event that the subject omitted an important point, he was referred once more to the instructions in question.

It will be noted that each condition of the experiment, with the exceptions of I, II, and III, presented more and more information relative to the nature of the experimental set-up—the expectation of recall, the expectation of the interpolated list, further knowledge concerning the possible effects of the interpolated list, and, finally, the suggestion to learn the material in such a way as to eliminate or at least to mitigate the effects of the interpolated list on the recall of the original list.

Condition III should furnish data showing what may be called 'raw' retroactive effects. Particularly is this true if the results of this condition are compared with those of Condition I where no interpolated learning was introduced into the experimental procedure. Then, comparing the findings under Condition III with those under Conditions IV, V, VI, and VII, we may possibly draw certain conclusions concerning the effect of attitude or set at time of learning on the disruptive effects called retroaction.

Of the two hundred subjects participating in this experiment, twenty-one served under both Conditions I and II; twenty under Condition III; thirty-three under Condition IV; thirty-four under Condition V; thirty-five under both Conditions VI and VII. All of these subjects came to the experimenter on four successive days.[4] On each of the first two experimental days, the subjects learned a list of nonsense syllables similar to the list of so-called original material to be learned on the third experimental day.

[4] In some instances, one day intervened between the first and second experimental days, or between the second and third experimental days; but at no time did more than twenty-four hours elapse between the third and fourth experimental days.

One of the objects of these two practice sittings was to familiarize the subjects with the experimental apparatus and procedure; the other, to provide the experimenter with some basis for equating the different groups of individuals. On the third experimental day, the subjects learned the nonsense material under the experimental conditions specified, and on the fourth experimental day, the recall was measured.

An average of the number of trials necessary in order to learn the first two lists was computed for every subject. Using such averages for the first sixty-eight persons participating in the experiment, the median, upper, and lower quartile points were computed. They were found to be as follows:

	Trials to learn
Median (Q_2)	30.75
Q_3	38.00
Q_1	25.25

The range of the average trials to learn was from 12–59.

Approximately the same number of poor, average, and good learners [5] were assigned to each group so that, under the several conditions, the groups are comparable as to the average number of trials necessary to learn the first two lists of nonsense syllables.

It may be argued that two practice trials do not afford sufficient basis for a classification of individuals as poor, average, and rapid learners. Possibly this argument is entirely valid, yet it seemed advisable to make some effort to equate the groups in as many respects as possible, *i.e.* age, sex, college year, and, lastly, learning ability.[6] Perhaps the only achievement was a rough classification within the groups of those individuals whom Meumann designated as 'die schnellen und die langsamen Lerner.'

[5] The poor learners were classified arbitrarily as those who learned in more than 38.00 trials; the average learned in from 25.26 to 38.00 trials; and the good learners were those who succeeded in 25.25 trials or less.

[6] It was suggested that the groups be equated as to intelligence or possibly high school record. The correlation between intelligence, as measured by the American Council on Education Test, and the average number of trials to learn a list of nonsense syllables was $r = +.18$; between high school rank and average number of trials, $r = -.03$. The first correlation above was computed on the basis of 185 cases; the second, on 203 cases. The decided lowness of these correlations influenced the writer against employing high school records or intelligence test scores in equating the groups. A larger number of practice trials in learning lists of syllables might have raised these correlations, but this seems questionable.

All of the learning throughout the experiment was done in the same room—a large quiet office—with only the experimenter present. No subject was asked to participate in the experimental work prior to 12 o'clock. In so far as possible, a subject worked at the same hour on the various experimental days.

The question of experimental controls also enters. When sets of directions are being given and their effects studied, and particularly when these directions are so arranged as to give more and more information concerning the experiment, it becomes of considerable importance that a subject, supposedly working under one set of directions or conditions, is not influenced by directions previously given to other subjects. In the present experiment, such a difficulty was obviated by completing each experimental condition, as far as the number of subjects was concerned, before proceeding to the subsequent condition.[7]

A phase of the experiment which may be open to criticism is the lack of control of the subjects by the experimenter dur-

[7] When it seemed advisable, in order to test the significance of various differences, to add to the number of cases under Conditions IV, V, VI, and VII, the rather troublesome situation arose of using subjects under Conditions IV and V, for example, whose acquaintances had, possibly, already worked under Conditions VI and VII. This difficulty was met in the following manner: Additional subjects under Conditions IV and VI were informed that the experimenter was working on an entirely different phase of the problem, one under which students had not previously served as subjects. Needless to say, subjects were added to Condition IV before additions were made to Condition VI. Condition V was held over until the Summer Session of 1930. At this time subjects of the same relative ages, and apparently unfamiliar with the experimental problem, from the elementary psychology course were employed. Condition VII, because it gave the most complete information, was not affected. The constancy of the average number of trials necessary to relearn the material and the number of syllables recalled during the first trial led to the conclusion that the additional subjects in Condition VI were unaffected by any other directions. The shift of means in the two measures of recall in Conditions IV and V, when added subjects had been employed, led the experimenter to run twenty-seven entering freshmen of the 1930–31 school year under these conditions—thirteen of these under Condition IV and fourteen under Condition V. The mean trials of these additional subjects to relearn and anticipate syllables resembled more the original twenty cases under each condition. This would seem to indicate that those cases in Conditions IV and V, added during the Spring and Summer Session of 1930, had been influenced by directions other than those given by the experimenter. For this reason, Conditions IV and V, as discussed here and in the following section, are each composed of the original twenty cases plus the added cases drawn from the entering freshman group of the college year 1930–31.

ing the twenty-four hours which elapsed between the learning of the original material and the recall. Although all subjects were directed not to practice on the lists after leaving the experimenter, there was no guarantee of their compliance with the request. In this regard, the present investigation suffers no more than any experiment in which absolute control of the subjects during the time interval between learning and recall is impossible or impracticable. Whenever directions stating that a recall will take place are given, any experimenter necessarily throws himself upon the weakness, or perhaps the strength, of his human material, and his findings must always be evaluated with that uncontrolled factor in mind. Of course an experimenter may attempt to insure accuracy by asking his subjects whether or not they have practiced on the lists of learned material, or he may utilize the more objective criterion of a recall performance which, by its obvious perfection, indicates additional practice. Such results would necessarily be discarded. However, it seems, in the main, practically necessary to rely on the integrity of the subjects themselves.

Section III. Results

Conditions I and II were introduced into the experiment merely as controls with the view, first, of ascertaining the effect of knowledge, or lack of knowledge, of recall on the retention of a previously learned list of nonsense syllables and, secondly, for the purpose of comparing retention in the presence and absence of interpolation of similar learning material. Table 1, given below, presents the retention in terms of number of trials to relearn and the number of syllables recalled the first trial for the first three conditions of the experiment:

TABLE 1

Average Trials to Relearn, and Average Number Syllables Recalled First Trial in Conditions I, II and III

Cond.	No. of Cases	Ave. No. Trials to Relearn List	P.E.	Ave. No. Syllables Recalled 1st Trial	P.E.
I	21	8.14	.540	4.24	.322
II	21	6.05	.421	4.57	.340
III	20	11.35	.525	.400	.132

The retention measured in trials to relearn in Condition II is significantly greater than the retention under Condition I, being 3.05 times the P.E. of the difference. At the same time, no significant difference exists between the conditions in the number of syllables recalled on the first trial. If one assumes that directions at time of learning,—in this case, knowledge of a subsequent recall,—establish a set or Einstellung that facilitates or energizes the learning process, such facilitation is not apparent in the amount actually recalled under Condition II. However, if one distinguishes as did Ebbinghaus between conscious retention and unconscious retention, then in Condition II there is evidence of greater unconscious retention. If the learning process and retention are thought of as involving basically a modification of neural processes, then a saving in trials necessary to relearn is a good index of the amount of permanent learning or acquisition. The amount of retention expressed by the savings score is 74.9 percent in Condition II and 66.6 percent in Condition I.

Contrasting Condition III with Conditions I and II, the disturbing effects of an interpolated list on the retention of previously learned material may easily be observed. As was to be expected, the differences in retention between Conditions II and III are greater than between Conditions I and III.

Although the present experimental procedure was not arranged to test the effect of the temporal position of the interpolated learning on the amount of retroactive inhibition, nevertheless, it incidentally brings out clearly the retroactive effects of such interpolated learning when it is introduced prior to recall. This evidence tends to refute Müller and Pilzecker's perseveration hypothesis as an explanation of retroactive inhibition.

Table 2 illustrates the retention in Condition III compared with retention in the other conditions of the experiment:

TABLE 2

AVERAGE TRIALS TO LEARN AND RELEARN, AND AVERAGE NUMBER SYLLABLES RECALLED FIRST TRIAL UNDER CONDITIONS I–VII OF EXPERIMENT

Cond.	No. of Cases	Ave. No. Trials to Learn List	Ave. No. Trials to Relearn List	P.E.	Savings Score	Ave. No. Syllables Recalled 1st Trial	P.E.	Percent of Material Recalled
I	21	24.33	8.14	.540	66.6%	4.24	.322	35.3
II	21	24.14	6.05	.421	79.4%	4.57	.340	38.1
III	20	25.30	11.35	.525	55.2%	.400	.132	3.3
IV	33	24.42	8.21	.338	66.4%	1.18	.180	9.8
V	34	24.32	8.79	.330	63.9%	1.23	.138	10.2
VI	35	22.91	7.48	.288	67.4%	1.54	.152	12.8
VII	35	23.48	6.83	.292	70.9%	3.08	.261	25.7

Condition III illustrates what has previously been referred to as the 'raw' effect of retroactive inhibition. In this condition, information, directions and mental set, in their relation to the phenomenon of retroaction, are at a minimum. This may, therefore, be considered the basic condition against which the other conditions involving retroaction are compared. The effectiveness of directions at the time of learning must be measured in terms of the greater amount of retention in those conditions where such directions are given as compared with the amount of retention in Condition III.

Condition IV is the first one in which directions are introduced along with the presentation of the interpolated learning. In this condition, the subjects were aware at the time of learning that a recall of the original material was to occur.

The directions given under Condition V led the subjects to expect a recall of the original list and also acquainted them with the fact that another list was to be learned prior to the recall.

The next set of directions—those employed under Condition VI—informed the subjects at the time of learning the original lists that a recall would take place, and that another list of similar syllables would be learned prior to recall. The possible interference effect of this list on the retention of the original learning was also stated.

In Condition VII, the subjects were provided with all of the information given in Condition VI, and the following statement was added: "To avoid mixing up the syllables of the

second list with those of the first list, take special care to keep them separate. Learn the first list in such a way that you will not confuse it with the second list."

A glance at the preceding table brings one to the obvious conclusion that set or attitude at time of learning, whether it be knowledge of recall, knowledge of the presentation of the interpolated material, knowledge of its possible effects, or an attempt to avoid the detrimental influence of the interpolation, has a significant effect on the amount of retention. This increase in the amount of retention holds true under both methods of measurement employed in this experiment.

A comparison of Conditions I and III, and II and IV respectively, shows definitely that retroaction is greater when measured in terms of recall than when tested by relearning. When interpolated learning is introduced, there is an average decrease in recall of 30.2 percent. If the retention measure is in terms of relearning, there is an average decline of 12.2 percent.

Having established experimentally that mental sets brought about by means of directions given at the time of learning do augment retention or, contrariwise, retard or reduce retroaction, we encounter the further question: Do the directions act differentially on the amount of retention? Does a knowledge of recall and expectation of an interpolated list increase retention more than a mere knowledge of recall? Further, does knowledge of the possible effects of an interpolated list increase retention more than just the bare expectation of such a list? And, lastly, will the directions, indicating that subjects should learn the original list in such a way as to prevent the interference effects of the interpolated list, be still more effective in increasing recall?

If, in arriving at conclusions, consideration is restricted to those differences between the means that are significant, the data seem to indicate little difference in retention between Conditions IV, V, and VI. Apparently, if subjects are not told to avoid the retroactive effect, *i.e.* to avoid interference resulting from the interpolated list, it is of little avail to notify them that there is to be such a list or to inform them of its

effect. In fact, mere anticipation of the presentation of a second list of similar material to be learned, or mere knowledge of the possible results therefrom seems no more effective in overcoming retroactive effects than knowledge of a recall of the original learning.

This conclusion applies to retention as measured by both criteria. Table 3, which follows, presents these differences between means in Conditions IV, V, and VI, the P.E. of the differences, and the number of times the differences exceed the P.E.

TABLE 3

DIFFERENCE BETWEEN MEANS, P.E. OF DIFFERENCE, AND TIMES DIFFERENCE EXCEEDS P.E. IN CONDITIONS IV, V, AND VI

Cond.	Difference Between Means		P.E. of Difference		Times Difference Exceeds P.E.	
	Trials to Relearn	Syllables Recalled First Trial	Trials to Relearn	Syllables Recalled First Trial	Trials to Relearn	Syllables Recalled First Trial
IV–V	.58(IV)	.05(V)	.473	.227	1.23	.02
IV–VI	.73(VI)	.36(VI)	.444	.236	1.65	1.53
V–VI	1.31(VI)	.31(VI)	.438	.205	2.99	1.51

Note: Numerals in parenthesis indicate condition favored.

However, if one follows the trends in the data presented by Table 3, it would appear that the results under Condition VI are consistently better than those under Condition IV or V. This result holds in both of the measures of retention. Further illustration of this tendency is apparent in the data given in Table 2. Comparing savings scores, Condition VI shows a retention of 67.4 percent as compared with 66.4 percent and 63.9 percent under Conditions IV and V respectively. If measured by the percent of material recalled, retention shows a progressive increase in amount as one proceeds from Condition IV to Condition VI. Thus, there appears to be a tendency for retention to be a function of the inclusiveness of directions.

Condition VI presented the opportunity for more motivation in the original learning, though this motivation was not established directly by instructions. When subjects are made

aware of the possible effects of a second learning situation, as in Condition VI, for some, if not all, of them there must be a self-induced set to counter-act, if possible, the disturbing effect.

The absence of any very significant differences between Conditions V and VI may be explained in part, at least, by assuming that the conditions were basically similar for some of the subjects under both conditions. Some subjects under Condition V, when told of the subsequent presentation of an interpolated list, inferred its effect. This was evidenced by expression of "Oh, that will mix me up!" Further evidence of the same reaction was obtained by asking a number of subjects, after they had completed the experiment, what ideas as to the effect of learning a similar list of syllables occurred to them upon reading the printed directions. They unhesitatingly replied, "Oh, I thought it would interfere."

The possibility also arises that those subjects in Condition V, who were not far-sighted or imaginative enough to foresee the possible consequences of the directions, benefited no more from knowledge of a recall and interpolated list, than did those subjects who knew only of a recall. This suggestion may account for the absence of any difference in retention between this Condition and Condition IV.

Any experimenter, at this point, finds himself confronted by those factors of individual differences which make the stimulus he controls, and believes to be somewhat objectively defined, a variable thing. A set of directions, stating that an interpolated list will be introduced prior to relearning, will produce in some of even a naïve group of subjects the thought, apprehension, or attitude of expected interference; in the case of other subjects, equally naïve as regards such concepts as retroactive inhibition and the like, no such outcome is envisaged. It appears to the present experimenter that such considerations may have eliminated or, at least, dimmed the supposedly clear-cut quantitative differences in directions. It is for this reason that the trends in favor of Conditions V and VI are significant. They seem to transcend the very likeness that may have come into the directions.

Turning now to a consideration of Condition VII, it will be remembered that subjects were here directed to learn in such a way that the second list would not be likely to interfere with the remembering of the first list. This attitude, induced by such directions, was presumably the most active. The following table shows Condition VII compared with Conditions IV, V, and VI.

TABLE 4

DIFFERENCE BETWEEN MEANS, P.E. OF DIFFERENCE, AND TIMES DIFFERENCE EXCEEDS P.E. IN CONDITIONS IV, V, VI AND VII

Cond.	Difference Between Means		P.E. of Difference		Times Difference Exceeds P.E.	
	Trials to Relearn	Syllables Recalled First Trial	Trials to Relearn	Syllables Recalled First Trial	Trials to Relearn	Syllables Recalled First Trial
IV–VII	1.37(VII)	1.90(VII)	.447	.317	3.09	6.00
V–VII	1.96(VII)	1.85(VII)	.441	.295	4.44	6.27
VI–VII	.65(VII)	1.54(VII)	.410	.302	1.58	5.09

Note: Numerals in parenthesis indicate condition favored.

All of the above differences in retention, with the exception of the difference between Conditions VI and VII, measured by trials to relearn, are significant, and show Condition VII to be the superior one.

The effectiveness of a more active set as represented by Condition VII at the time of learning, in relation to retroactive inhibition may possibly be more vividly illustrated by a comparison with retention in Conditions I and II. Measured by the number of trials to relearn, Condition VII is superior to a condition in which no retroactive effects occur; and only slightly inferior to that in which a knowledge of recall and no retroactive effect occurs. Computing the savings score, there is a retention of 70.9 percent in Condition VII; 66.6 in Condition I; and 79.4 in Condition II. In terms of the percentage of material actually recalled, the retention is 25.7 for Condition VII; 35.3 for Condition I and 38.1 for Condition II. This latter finding also indicates that the effects of retroactive inhibition are greater on the phase of retention which is measured by the recall method.

Section IV. Conclusions and Discussion of Results

Conclusions

The foregoing experimental project has brought to light several findings worthy of summary and discussion.

1. Directions to the subject informing him what to anticipate at the time of learning,—*i.e.* presence of a recall; expectation of interpolated list; information in regard to the possible effects of an interpolated list—any one or all of these presented at the time of the original learning very definitely minimize the retroactive effects and increase the retention of the original material.

2. There is a consistent trend in the data pointing to increasing degrees of retention with increasing amounts of information given at the time of learning, although the differences, taken individually, do not prove to be statistically significant.

3. Directions urging the subject to make an effort to avoid the interference effects of the interpolated list result in a significantly greater retention than is obtained under any of the other conditions involving informatory directions.

4. The presence of considerable retroactive inhibition has been shown to exist when the interpolated learning of similar nonsense material is introduced immediately prior to a recall of the originally learned nonsense material and twenty-four hours after the original learning.

Such a finding, while in no sense new, would seem to further substantiate Robinson's (**5**) conclusion that retroactive inhibition is not due to the disturbance of a perseverative process immediately after learning, but is a function rather of some other factor, such as associative interference, due to the similarity between the interpolated activity and the original learning.

It might be interesting to duplicate this same experiment in all respects except to introduce the interpolated learning immediately after the original learning. This would decide the question as to whether interference with any hypothetical perseverative tendencies would in any way change the results here obtained.

5. Retroactive effects are definitely more detrimental to

retention as measured by the recall method, than to retention measured by the relearning method.

The findings of the present investigator corroborate in this respect the results of McGeoch (**3**), and are somewhat similar to the early experimental conclusions of Heine (**2**). However, the latter worked with recognition and recall as measures of retention, and hence it is only by considering recognition and relearning as being somewhat alike as measures of retention, in contrast to recall as a measure, that we can compare the two results.

Discussion

The main outcome of the experiment is, therefore, its demonstration that directions establishing a set or attitude at the time of learning have the effect of minimizing or curtailing retroactive inhibition. Some explanation of this decrease in so well recognized a phenomenon is necessary.

The word 'set' may be variously defined. For instance, it may denote a bodily posture temporarily assumed by an organism, *i.e.* the runner toeing the line, or the cat crouching for a mouse. Here is an illustration of a preparation for action which involves observable changes in the adjustment of muscles. But set may also be defined as that mental attitude which is the result of specific work instructions, and which may or may not involve changes in muscular tonus, or the assumption of any particular bodily attitude, on the subject's part. It is this latter definition or meaning of set with which we are dealing in this study. It seems probable that the directions, acting as a stimulus, established an integrated neural pattern which energized or facilitated the learning process that was being carried on.

As Bills (**1**) pointed out, the dynamogenic influence of simultaneous processes on one another has long been recognized, although the analysis of these processes and of the phenomenon of facilitation has not been investigated very adequately, either on the subjective or on the neurophysiological side. A theoretical explanation is offered in the drainage hypothesis of McDougall which, in rather vague terms, attempts to account for facilitation by assuming that an on-

going activity, such as learning, drafts the neural energy from impulses simultaneously aroused in adjacent neural systems. When we talk in terms of set we have not advanced very far from the Einstellung of Watt and the determining tendency of Ach and Woodworth.

Inherent or implicit in the concept of determining tendency is the implication that the tendency operates superiorly to reinforce certain reactions, *i.e.* a determining tendency to add or subtract or multiply in the case of $\frac{5}{2}$ strengthens one association—7, 3, or 10 respectively—and the implications are that such a set would operate in the formation of such associations. Peterson (4), in a study on the effect on learning of attitudes produced by the knowledge or lack of knowledge that recall would be required, obtained differential results for the two attitudes. He refers to the 'will to learn' attitude, as discussed by Ebert and Meumann, and the fact that this attitude exerts an effect on the efficiency of memory. He states, furthermore, that particular adjustments of this attitude are manifested, *i.e.* if the subject endeavored to recall meanings rather than words, a better reproduction of meanings occurred and vice versa. In reference to his own experimental data, Peterson contends that the additional stimulus which produced the will to learn was the statement to students, before giving the list, that they would be asked to reproduce the words contained therein. This knowledge, Peterson thinks, produced, through various neural dispositions, a far more definite response attitude than was possible when no directions were given. It may be that our directions operated in a similar fashion.

Mention has already been made, in the section devoted to the results of this experiment, of the greater retention, measured by the savings method, in Condition II than in Condition I. Condition II gave knowledge of a recall. Here the attitude at the time of learning apparently facilitated the learning —in the sense of bringing about a more complete acquisition. It will be noted that the average number of trials to learn was approximately the same, in each case.

In order that more definite proof might be obtained of the

energizing effect of directions on learning and on subsequent retention, eleven subjects were given directions similar to those under Condition VII and on the fourth experimental day were asked only for a recall. The condition thus obtained involved a more active learning attitude than any other condition in the entire experimental procedure, with the exception of Condition VII. The average number of trials necessary to learn the original list was 23.54; the number of trials to relearn, 4.30. Expressed in terms of savings score there is a retention of 81.5 percent. This amount of retention is rendered more conspicuous when compared with the retention of 66.6 percent and 74.9 percent under Conditions I and II.

There is also some indication of the facilitating effect of directions, apparent in the average number of trials required to learn the original lists under the several conditions of the experiment. Assuming, as seems justifiable, that our groups were approximately equal in respect to the capacity or ability for learning meaningless material, the tendency for the average trials to learn, to decrease as the directions become more complete would seem to indicate the greater effectiveness of such directions in energizing the learning. Reference to Table 2 in the section on results will make more clear the above inference.

If we assume that the interpolation of similar learning material introduces a disruptive force affecting the retention of the original learning, the natural expectation would be that the more efficiently learned material would be more easily retained under interference conditions. It would also be anticipated that the subjects possessing some expectation of the requirement of a recall, and of the introduction of interpolated material, would be less influenced by the actual presentation of the latter. In other words, preparation for a future contingency greatly lessens the startling or upsetting effect of its actuality.

Reference to Table 2 will illustrate that retroactive effects are more apparent when the number of syllables recalled is taken as a measure of retention. The differences between the several conditions of the experiment are also greater when

this criterion is employed. This is to be explained largely in terms of the interference which occurs when a recall is demanded after the learning of the interpolated material. Considerable confusion is evident in the lists as recalled at this time. For instance, the transfer causing retroaction may be, in one case, a shift from original memorizing to interpolation; and, in another, a transfer from interpolated to original material. Such interference would make itself apparent during the subject's immediate recall more than in his trials to relearn since the exposure of the original list and the subsequent presentations allow an opportunity for the interference effects to wear off.

In our discussion of the active attitude or set, we have assumed that it acted directly upon the learning process to energize it. But there is an alternative possibility that the effect of set or attitude on retention could be explained on the basis of a theory of perseveration. Although the present experimental procedure, since it entails a lapse of twenty-four hours before the recall, does not preclude this hypothesis, the writer feels such a theory to be highly improbable.[8] It might be suggested that subjects under the directions of a recall, with knowledge of an impending interpolated list, or with an active set to overcome the interference effect of such a list, were put in a condition (mild anxiety, certain amount of tension, desire for a good performance) that favored the perseverating, and 'setting in' of these lists of nonsense syllables. That is, the directions created a background of mental events which might have augmented the perseverative tendencies. Such a conclusion is not entirely inconsistent with material already at hand concerning a tendency to greater perseveration under conditions of anxiety, worry or remorse. Watt's theory (7) of reluctancy fits into this interpretation also.

Although the field of neural organization is still relatively uninvestigated, explanations eventually should be forthcoming; and the present experiment may be considered as a small offering along the line of descriptive evidence on the effect of mental set in relation to retroaction.

[8] It is hoped that future experiments will obtain definite and conclusive data on this point.

The words of Watt are still a challenge:

"If we wish individuals to respond to a series of single words in a specified way and they can do this, and firmly resolve to do it, we shall find that no number of repetitions will make strong enough the associations between the words of the series and other words, so as to bring the wrong words out against the intention of such individuals."

(Manuscript received April 7, 1932)

References

1. Bills, A. G., The influence of muscular tension on the efficiency of mental work, *Amer. J. Psychol.*, 1927, **38**, 227–251.
2. Heine, R., Über Wiederkennen und rückwirkende Hemmung, *Zsch. f. Psychol.*, 1914, **68**, 161–236.
3. McGeoch, J. A., Influence of degree of learning upon retroactive inhibition, *Amer. J. Psychol.*, 1929, **41**, 252–262.
4. Peterson, J., The effect of attitude on immediate and delayed reproduction, *J. Educ. Psychol.*, 1916, **7**, 523–532.
5. Robinson, E. S., Some factors determining the degree of retroactive inhibition, *Psychol. Monog.*, 1920, **27**.
6. Watt, H. J., *The Economy and Training of Memory*, Longmans, Green & Co., 1909.
7. Watt, H. J., *The Common Sense of Dreams*, Clark University Press, 1929.

HABITUATION AND PERSEVERATIONAL CHARACTERISTICS OF TWO FORMS OF INDIRECT SUGGESTION

BY CLARK L. HULL AND MILTON C. FORSTER
Yale University

I

Whether two distinguishable natural phenomena are essentially alike or different is always a nice question. Queries concerning essences usually have little meaning and are capable of no very significant answer unless the point of view of the questioner is known. The prime object of the present investigation is to clarify certain ambiguities concerning the nature of suggestion as related to hypnosis. Various phenomena such as hypnosis, direct waking suggestion to movement, and the forms of indirect suggestion represented by Binet's progressive weights and his progressive lines have long been classed without distinction as suggestion. Are all these phenomena essentially allied to hypnosis, or are some of them fundamentally unrelated to it? Experimental evidence is accumulating which raises some presumption in favor of the view that hypnosis and direct waking suggestion to movement are of the same general nature in the sense that the behavior of the two seems to conform to the same basic pattern. It has been shown by Krueger (4), for example, that falling into the hypnotic trance as indicated by the rate of eye closure is positively facilitated by repetition or practice. Hull and Huse (3), Williams (7), and especially Patten, Switzer, and Hull (5) have shown that direct waking suggestion is also positively facilitated by repetition. On the other hand, certain *a priori* considerations lead to the view that the progressive-increment forms of indirect suggestibility are essentially distinct from hypnosis and that one among several significant distinguishing characteristics might well be that practice or repetition should negatively influence susceptibility to the

tendency in question. Some observations made by Williams (7, 40 ff.), in the course of an experiment set up primarily to solve an entirely different problem, tend more or less to corroborate this hypothesis. If this distinction can be definitely established it will tend considerably to simplify the problem of defining suggestion (8, 78) and will ultimately aid in the clarification of the relationship of suggestion to hypnosis, at present in a distinctly unsatisfactory state.

II

It was believed that in the present investigation clear-cut results would be more readily obtained with subjects who showed an intermediate amount of suggestibility, since subjects showing complete resistance to suggestion could not display any increase in resistance as the result of repetition, and subjects showing complete positive response could not possibly manifest any change in the direction of increased susceptibility through practice. In a preliminary experiment with university students it was found that under the conditions of the present experiment they were so slightly responsive to the forms of suggestion employed that they were practically useless as subjects. It has long been known, however, that children are far more responsive to all kinds of suggestions than are adults. Accordingly, the subjects finally chosen for the present investigation were boys between the ages of 10 and 11 years. Because they were living in an orphanage they were readily available for experimental purposes. Fortunately the greater part of the subjects had received intelligence tests. These scores, supplemented by the careful observations of the experimenter, indicated that those children finally chosen for the experiment were entirely normal individuals.

The forms of suggestion chosen for investigation were the classical ones originally devised by Binet (1), the two based respectively on progressive weights and progressive lines. With minor exceptions the apparatus and technique of administration were the same as those employed by Williams (7, 17 ff.). The fifteen weights were made from rectangular pill boxes, the first four weighing 20, 40, 60, and 80 gr., and the remaining eleven weighing 100 gr. each. The boxes were

numbered beneath, but all appeared to be exactly alike to the subject's view. They were always arranged in the same order on a ledge which was at the height of an ordinary table.

The subjects were given oral instructions at the beginning of each day's experiment, as follows:

> I want you to lift this weight (pointing to the first weight) and then this weight (pointing to the second weight) and tell me whether this one here is heavier or lighter or just the same weight as that one (pointing again to the first weight). Then lift this one (pointing to the third weight) and tell me whether this one is heavier or lighter or just the same weight as that one (pointing to the second weight) and then do the same for the rest of these.

In order to make as certain as possible that the subjects fully understood the task, a preliminary practice series was given in which the weights commonly used in the Binet intelligence tests were used. Practice with these weights arranged in different orders was continued until it appeared clear that the subject was really reacting to the differences in weight involved rather than merely guessing at random. Several subjects who appeared to react carelessly or automatically were rejected. Each subject in the experimental group made the series of fourteen judgments in the experiment proper three times at the same hour on each of ten consecutive days. Nine subjects made the three series in immediate succession, but with five, each series was separated from the next by approximately six minutes of play. At the conclusion of each series of judgments the subjects were encouraged by the experimenter, who would say, "That is good! Now do it again," or, "That is fine! Now do it again and do your very best this time," and the like.

The apparatus employed in the parallel experiment of the progressive lines was the one used by Williams (7, 17). It consisted of a board screen in which was cut a rectangular hole 2 cm high and 11.5 cm wide. Behind the screen was an attachment for mounting a cardboard disk upon which were placed at intervals fifteen lines drawn in India ink. The lengths of the first four lines were 12, 24, 36, and 48 mm, and the remaining eleven, 60 mm each. The disk could be revolved so that the lines would appear at the window one at a time, in the order given above. The lines were presented

uniformly 1 cm from the bottom of the aperture, but to discourage possible tendencies to stereotyped reaction the lateral positions of the lines as they appeared in the aperture were systematically varied. It is well known, however, that the setting in which a line is viewed greatly influences its apparent length. To equate this factor of position, nine different disks were drawn up in such a way that each line of the series would appear in each lateral position at the window an equal number of times. The disks were systematically rotated among the several subjects in such a way that for the group as a whole no constant error due to position at the window would enter the judgments of length at any point in the judgment series.

The instructions to the subjects in this portion of the experiment were also given orally, as follows:

> I am going to show you some lines and I want you to tell me if the second line that you see is longer or shorter or just the same length as the first line that you saw. Then I'll show you another line, and you tell me whether that one is longer or shorter or just the same length as the line you saw just before it.

A preliminary practice series was given to make sure that the subject understood the nature of his task. In the experiment proper three series of judgments were given each subject each day, and the experiment was continued for ten consecutive days, exactly paralleling the procedure with the weights. A different sequence of disks was employed on each day and with each subject. The same type of encouragement was given these subjects as those who served in the weights project.

The experimental work of the entire investigation was conducted in a small room at the orphanage. Each child was called into the experimental room and made his judgments individually. While awaiting their turn to serve in the experiment, the children amused themselves in an adjoining room with puzzles and games. From such information as the experimenter was able to gather, the children appear to have been so much engrossed with their amusements that they indulged in very little or no conversation concerning the experiment. Naturally they were given no hints as to the nature of the problem under investigation.

No child acted as subject on more than one form of suggestion.

III

In both forms of suggestion under consideration any normal subject will judge the second, third, fourth, and fifth presentations as greater than the immediately preceding one, for the reason that they are obviously so in fact. As the result of this preliminary activity, there persists over into the remainder of the series a tendency to judge each succeeding presentation as 'heavier' (or 'longer'), even though they really are all alike. In this lies the evidence of suggestibility. Actually, however, the subject frequently also responds with 'same' or 'lighter.' Accordingly, the results of fourteen subjects who completed the weights experiment are presented in Table I from all three points of view. An inspection of the row of means shows that there is a very marked and consistent tendency for the judgments of 'heavier' among the 100-gram weights to grow progressively less with increasing practice. This constitutes a preliminary indication of a decided negative practice effect. There is evident also a parallel tendency for the number of 'lighter' and 'same' judgments to become greater.

It is common in determining the extent of suggestibility in such experiments to consider only the judgments of 'heavier' and to ignore more or less completely the other two types of responses (**2**, 341). It is clear, however, that a subject might give a fairly appreciable score of 'heavier' judgments and yet have no tendency to suggestion whatever. He might, for example, on a given occasion judge 33 per cent 'heavier,' 33 per cent 'lighter,' and 34 per cent 'same.' In such a case there is just as much reason to conclude that there is suggestion in the direction of judging the weights too light as to judge them too heavy. As a matter of fact, such a situation presents no evidence of suggestion of any kind. Yet a naive interpretation of the 'heavier' scores considered alone might lead to the completely erroneous belief that a very appreciable amount of suggestibility is present. It is only as the three types of judgments as a whole are shifted to one side or the other from a symmetrical arrangement that any evidence of suggestion can appear. By the same token, it is evident that the amount

TABLE I

SUMMARY OF THE TOTAL NUMBER OF JUDGMENTS OF EACH TYPE MADE ON THE THREE TRIALS OF EACH DAY AT THE PRESENTATION OF WEIGHTS 6 TO 15 IN THE PROGRESSIVE-WEIGHTS EXPERIMENT

The first nine of these subjects performed the three trials in immediate succession. With the remaining five the trials were separated by play intervals of approximately six minutes each.

Subject	'Heavier,' by days										'Same,' by days										'Lighter,' by days									
	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
T.C.	12	7	6	9	4	8	12	7	11	8	18	23	23	21	26	22	18	23	19	22	0	0	1	0	0	0	0	0	0	0
C.I.	12	11	7	10	8	11	14	4	6	3	15	14	13	13	19	18	9	24	22	26	3	5	10	7	3	1	7	2	2	1
A.P.	28	29	29	21	24	21	20	17	20	24	2	1	1	9	6	9	10	6	7	5	0	0	0	0	0	0	0	7	3	1
H.S.	29	29	30	30	23	24	14	4	10	9	1	1	0	0	6	4	15	26	20	21	0	0	0	0	1	2	1	0	0	0
R.V.	20	17	11	6	3	3	1	4	3	2	9	11	18	23	26	27	27	24	25	27	1	2	1	1	1	0	2	2	2	1
J.K.	30	18	16	13	14	18	18	15	8	8	0	9	13	14	10	7	9	15	22	22	0	3	1	3	6	5	3	0	0	0
R.M.	20	22	22	20	21	21	17	24	18	15	0	0	0	0	0	0	0	0	0	1	10	8	8	10	9	9	13	6	12	14
F.M.	30	30	30	21	23	18	10	5	6	6	0	0	0	9	7	12	20	25	24	23	0	0	0	0	0	0	0	0	0	1
A.C.	16	17	20	18	11	9	7	11	6	12	9	5	1	2	4	10	8	8	10	5	5	8	9	10	15	11	15	11	14	13
E.B.	21	24	26	26	26	25	25	24	26	29	2	0	0	0	0	2	0	0	0	0	7	6	4	4	4	3	5	6	4	1
F.C.	28	19	20	20	10	13	10	17	15	15	0	1	2	2	5	7	7	6	5	7	2	10	8	8	15	10	13	7	10	8
H.B.	28	30	18	17	12	7	10	6	5	8	0	0	10	7	9	7	13	15	14	11	2	0	2	6	9	16	7	9	11	11
A.S.	15	15	14	10	11	10	12	17	13	14	3	6	8	9	7	6	1	0	2	1	12	9	8	11	12	14	17	13	15	15
C.B.	30	30	30	30	30	21	19	16	17	16	0	0	0	0	0	9	11	14	10	10	0	0	0	0	0	0	0	0	3	4
Mean per trial	7.6	7.1	6.6	6.0	5.2	5.0	4.5	4.1	3.9	4.0	1.4	1.7	2.1	2.6	3.0	3.3	3.5	4.4	4.3	4.3	1.0	1.2	1.2	1.4	1.8	1.7	2.0	1.5	1.8	1.7

of the tendency in question is measured by the magnitude of this shift.

These considerations led to the development of a method for determining from such judgment data the amount of suggestibility. The method proceeds upon the assumption, usually made in treating reactions of this kind, that all the judgments taken together are distributed according to the normal 'curve of error.' A convenient table of this function in terms of the standard deviation may be found in Thorndike's *Mental and Social Measurements* (**6**, 198). The method may be illustrated by the computations for determining the composite suggestion tendency on day one. Table I shows that this group averaged on this day 76 per cent 'heavier' judgments, 14 per cent 'same,' and 10 per cent 'lighter.' Since the middle of the area of 'same' judgments should coincide with the middle of the distribution of judgments as a whole if there were no distortion, the extent of a distortion is shown by the distance from the one to the other in cases where they do not coincide. The 76 per cent 'heavier' judgments occupies the entire 50 per cent on the right wing of the distribution (Fig. 1) and also the inner 26 per cent of the left wing. Reference to the table shows that 26 per cent of the statistical population occupies a space in this region of the distribution corresponding to .71 σ on the base line. Next there is added the 14 per cent of 'same' judgments to the 26 per cent of 'heavier' already on the inner side of this wing, making a total of 40 per cent. Reference to the table shows that 40 per cent occupies a space from the center outward corresponding to 1.28 σ on the base line. Thus one margin of the 'same' judgments falls at .71 σ, and the other at 1.28 σ, both on the left wing of the distribution. The range of the 'same' judgments is consequently the difference between 1.28 and .71, or .57 σ. The mid-point of the 'same' judgments is therefore one-half of .57, or .28 σ distant from the margin of the heavier judgments. But since the margin of the heavier judgments is .71 σ from the center of the distribution (the normal position for the mid-point of the 'same' judgments) the latter is .71 σ + .28 σ, or .99 σ from its normal position. Accordingly,

.99 σ is the extent of the displacement of the mid-point of the 'same' judgments, in this case presumably due in the main to the indirect suggestion from the preceding series of weights which really were progressively heavier. This outcome of the computation shows, then, that on the first day of practice there was a very considerable suggestion effect.

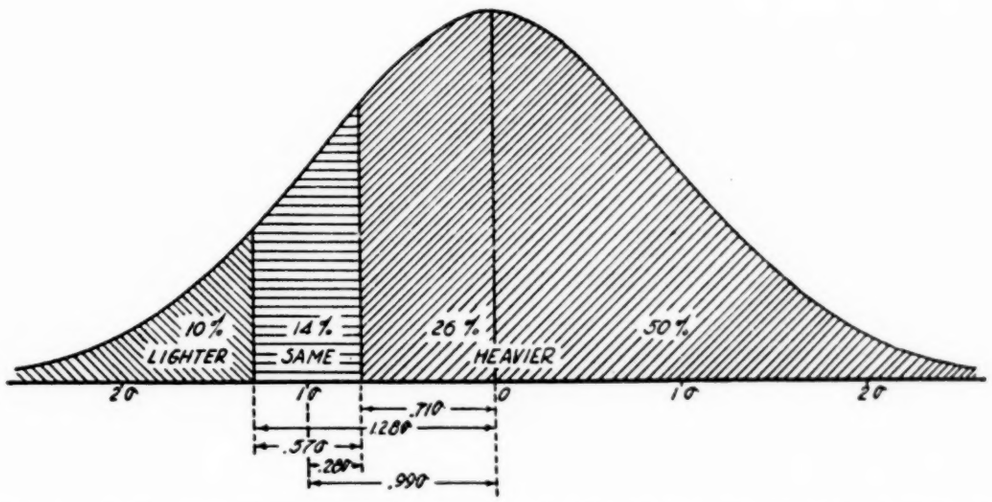


FIG. 1. A diagram illustrating the method of determining the amount of distortion in judgment of lifted weights due to suggestion. See text for explanation.

Similar computations were made for the remaining nine days of practice. These values are shown graphically in Fig. 2. An inspection of this graph reveals a typical learning curve with the usual tendency to flatten out as practice continues. As anticipated, the influence of repetition weakens the tendency to suggestion rather than strengthens it. During the ten-day period the rather strong tendency of the first day is gradually reduced until on the last day it is only one-third of its original strength.[1]

Unfortunately only three subjects completed the lines experiment by the technique paralleling that of the weights. The results from these individuals are summarized in Table

[1] This ignores the possibility that other factors may be operative here. It is possible, for example, that the well-known tendency to call the second of a pair of lifted weights heavier, observed in the classical experiments concerned with Weber's law, may be partly responsible for the distortion of judgment still present on the tenth day. This problem is capable of straightforward solution, but the limited purpose of the present investigation does not require it.

II. While such a limited quantity of data does not justify the statistical analysis given the progressive-weights results, an inspection of this table reveals some indications of a nega-

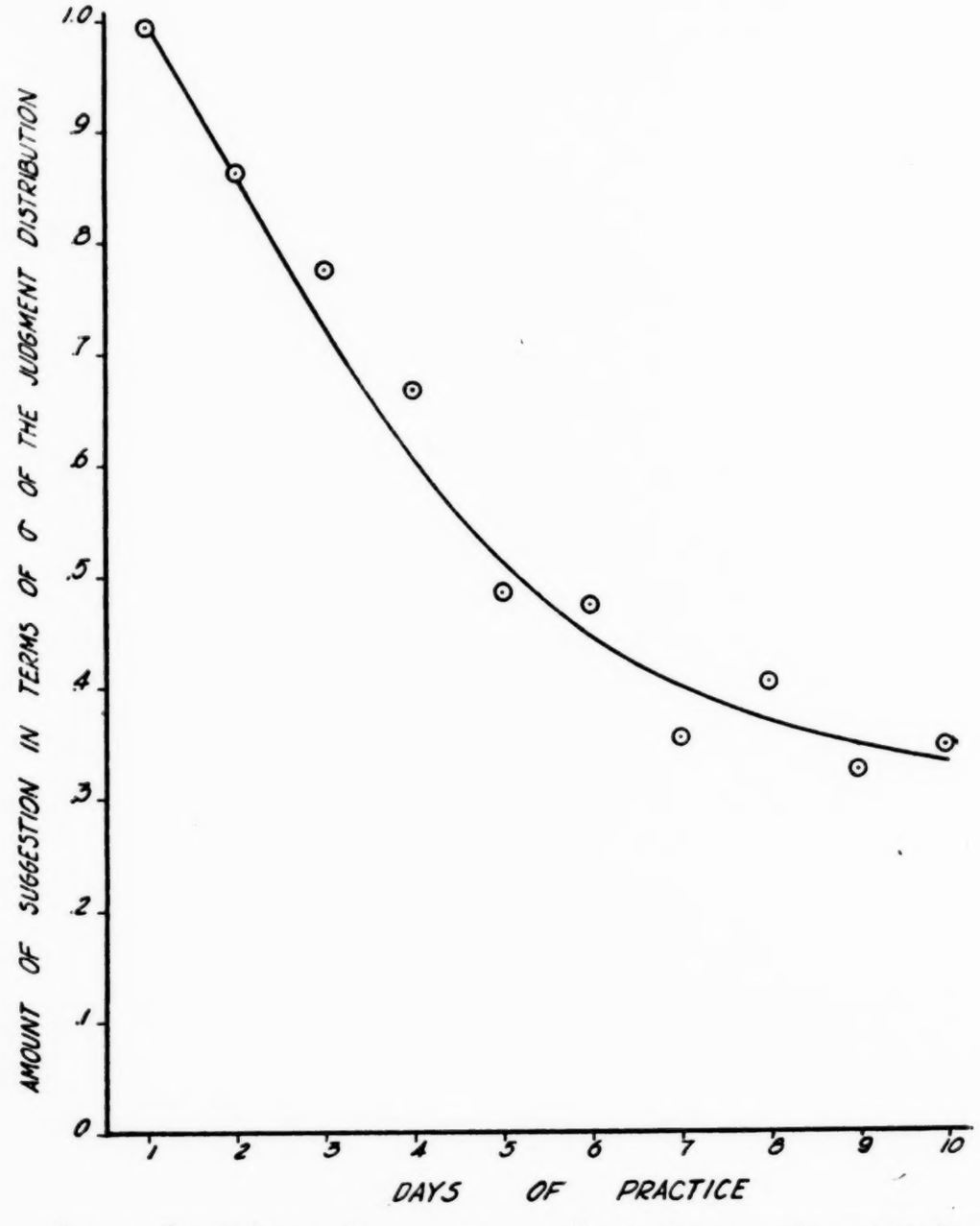


FIG. 2. Composite practice curve in responding to the progressive-weights form of indirect suggestion, in terms of the σ of the distribution of judgments. Note the decided negative practice effect.

tive practice effect. Final decision on this matter must, however, await a more adequate volume of experimental results.

TABLE II

SUMMARY OF THE MEAN SCORES OF THREE SUBJECTS 10 TO 11 YEARS OF AGE ON THE PROGRESSIVE LINES EXPERIMENT, WITH ORAL RESPONSES

No. of day	1	2	3	4	5	6	7	8	9	10
Mean no. of 'longer'	5.0	4.1	3.8	3.1	2.8	2.6	2.2	2.7	2.2	2.8
Mean no. of 'same'	2.8	4.0	4.0	4.4	5.4	6.3	7.0	5.8	6.8	6.0
Mean no. of 'shorter'	2.2	1.9	2.2	2.4	1.8	1.1	.8	1.6	1.0	1.2

IV

It happens that the form of the experiment with the lines permits of another and entirely different technique of recording the subjects' judgments, which throws some light on the nature of the suggestion process. Ten subjects were asked, instead of making a comparison of the successive pairs of lines with oral responses, merely to draw on a piece of ruled paper with a pencil a line equal to each as presented in the aperture of the screen. The lines drawn from one disk were removed before those on the next disk were presented. The instructions to the subject were:

> I am going to show you some lines and I want you to draw a line here which is just as long as the line you see.

Otherwise the experiment was conducted exactly as that described above in which oral reactions were required.

In view of the decided negative practice effects shown in the progressive weights experiment it might perhaps be expected that the mean length of the lines 5 to 15 drawn from day to day would gradually become less and less. Actually, no such tendency was found. The mean length of these lines in mm for the ten days was as follows: 54.1, 54.5, 56.0, 56.7, 55.1, 54.6, 53.8, 55.1, 55.3, 54.9.

In an attempt to secure some light on this paradoxical situation, we converted the line measurements into scores presumably equivalent in some sense to the scores from oral reactions to the same stimuli. If the second of two successive lines was drawn longer than the first by more than 4 mm it was scored 'longer,' if shorter by more than this amount, it was scored 'shorter.' If they did not differ by more than this amount they were scored 'same,' on the assumption that they

would not have been discriminated. The results of this procedure are summarized in Table III. An inspection of the row of longer values, if considered alone, might easily be mistaken as showing a distinct negative practice effect. A careful statistical examination of these converted scores by methods already described for oral judgments failed to reveal a practice effect of any kind. The indication seems to be that this form of progressive lines suggestion is very slightly, or not at all, influenced by a very considerable amount of repetition.

TABLE III

MEAN SCORES FOR TEN SUBJECTS 10 TO 11 YEARS OF AGE ON THE PROGRESSIVE-LINES EXPERIMENT WHERE THE LINES DRAWN BY THE SUBJECTS WERE MEASURED AND THEN CONVERTED INTO THE EQUIVALENTS OF VERBAL REACTIONS, ASSUMING THE THRESHOLD OF DISCRIMINATION TO HAVE BEEN 4 MM.

No. of day	1	2	3	4	5	6	7	8	9	10
Mean no. of 'longer'.....	3.90	3.47	3.40	3.23	3.23	2.77	2.70	3.03	2.97	2.53
Mean no. of 'same'.....	3.40	4.27	5.13	5.27	5.53	5.90	6.10	5.97	5.83	6.60
Mean no. of 'shorter'.....	2.70	2.27	1.43	1.50	1.23	1.40	1.20	1.00	1.17	.90

V

We may next inquire what evidences the two forms of indirect suggestion present of perseverational tendencies. Pursuing the analogy of the Patten, Switzer, and Hull experiment (**5**), we will first examine the relation of the strengths of the suggestion tendencies on the three tests per day throughout the practice series. Too few subjects were included in the progressive-lines experiment by the oral-response method to make profitable an examination of these data for this purpose. The statistically treated results of the judgments of the fourteen subjects in the progressive-weights experiment are summarized, from this point of view, in Table IV.

An inspection of the main body of this table shows that if any tendency to perseveration from one series of judgments to the next exists it is exceedingly slight and unstable—the means for the three trials interlace so much that it would be difficult to distinguish one from the other if they were to be plotted as were the results of the direct suggestion by Patten,

TABLE IV

SUMMARY OF THE COMPOSITE AMOUNT OF SUGGESTIBILITY OF 14 BOYS BY DAYS ON THE PROGRESSIVE-WEIGHTS EXPERIMENT FOR THE THREE RESPECTIVE TRIALS EACH DAY

The unit of measurement is the standard deviation of the judgment distribution.

Day	1	2	3	4	5	6	7	8	9	10	Means days 1–5	Means days 6–10	Means days 1–10
First trial.....	.87	.91	.83	.66	.43	.47	.39	.36	.34	.31	.74	.37	.56
Second trial....	.97	.85	.69	.52	.61	.45	.26	.33	.40	.38	.73	.36	.55
Third trial.....	1.18	.84	.88	.81	.40	.52	.44	.54	.23	.42	.82	.43	.63

Switzer, and Hull. Something roughly approximating this, however, may be seen at the right of the table where the pooled results of the first and second five-day periods of the practice are given. There emerges here a certain consistency, though the second trial shows a slight shift in a direction opposite that of the third. The amount of the tendency to perseveration is slight, only .07σ where the initial score of the day was .73σ or less than 10 per cent. It may be noted, incidentally, that this change within a day's practice is in the direction of an increase in suggestibility. This is opposed to the tendency of change from one day to the next, which is in the direction of a decrease. This means that none of the mean (negative) practice effect in this function takes place during the practice itself, but seems, somehow, to take place during the twenty-four-hour interval between practice periods.

VI

The form of the progressive-increment suggestion experiments permits of a second and fairly distinct quantitative approach to the phenomenon of perseveration. As already pointed out, the basic mechanisms of the suggestion responses in these cases presumably is the tendency of the organism which has reacted in a certain way to a series of gross differences to persist or perseverate into a stimulus situation where these reactions are no longer appropriate. But since the series goes on for a considerable distance after the termination of the exciting cause, it might be anticipated that the tendency to error would be at a maximum at the sixth presentation

and that it might gradually fade out as the successive judgments were made. In that event, a graph representing the magnitude of the suggestion effect through the portion of the series where the presentations were actually equal ought to represent the course of the perseverational tendency. Such a graph for the fourteen subjects of the progressive-weights experiment is shown in Fig. 3. Superficially the persevera-

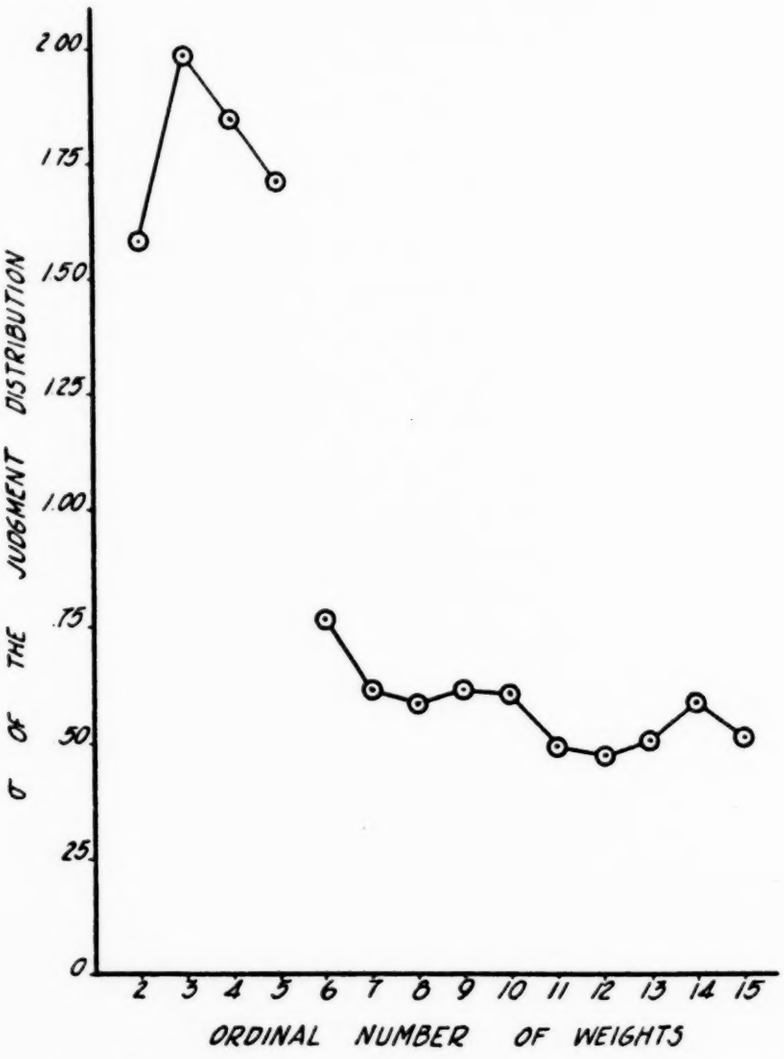


FIG. 3. Graph showing the composite tendency to judge the several weights of the series heavier than the one preceding. The fading out of the perseverational tendency is shown by the gradual approach to the horizontal of the graph for lines 6 to 15.

tional tendency is represented (for weights 6 to 15) by the distance from the graph to the base line. This tendency is nearly half as great as that on weights 2 to 5, where differences

of 20 gm. actually exist. The gradual approach of this curve to the horizontal shows the progressive fading out of the perseverational tendency. As already pointed out, it is probable that a portion of what appears to be suggestion here in reality is due to a tendency to call the second of a pair of weights heavier regardless of suggestion.

A still different view of perseveration may be obtained by plotting the composite curve of the lengths of the fifteen lines as drawn by the ten subjects on each of the ten days of practice. Such a graph is presented in Fig. 4, where it may be

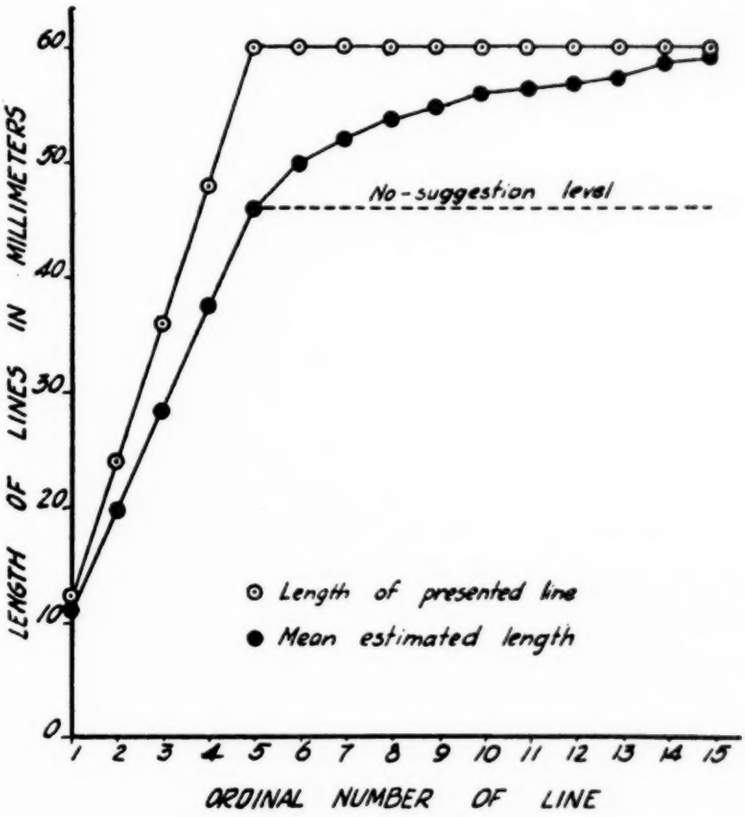


FIG. 4. A graph representing the perseveration of the composite tendency to draw successive lines longer after the lines to be reproduced no longer increase. The first five lines presented were consistently underestimated. If the remaining lines had been underestimated in the same way, lines 6 to 15 would have fallen along the broken line labeled 'no-suggestion level.' The tendency to increase the length of the lines gradually fades out with successive trials.

compared with the lengths of the lines as actually presented. There it may be seen that though the subjects as a group tended to underestimate the lines presented, their drawings increase in length with remarkable uniformity up to the fifth.

At this point, if there were no influence of suggestion, the curve presumably should have become horizontal, *i.e.*, it should have followed the course shown by the broken line. Instead we observe that the sixth line is increased nearly half as much over the fifth as the fifth is over the fourth—surely a very considerable influence of perseveration. In general the tendency to increase the lines persists through the entire series of ten reactions, though, as might have been expected, the tendency grows progressively weaker with the increased distance from its source, just as was observed in the case of the progressive weights. The remarkable smoothness and uniformity of this graph is distinctly convincing as to the genuineness of the tendency and its point of origin.

VII

In summarizing the results of the present experiment it may be said that the two forms of indirect suggestion investigated definitely differ from the forms of hypnosis and direct waking suggestion so far examined, in not showing a positive practice effect. The progressive weights display a very marked practice effect, but in a negative direction. It is possible that the suggestion responses to the progressive lines by the oral technique also are influenced in a negative manner by repetition, but the quantity of experimental data available is not adequate to permit of an opinion. The progressive lines by the reproduction technique appear to be nearly, if not quite, destitute of any practice effects in either direction.

In respect to perseveration effects, it may be said that the progressive weights manifest a slight tendency to yield greater suggestibility from trial to trial. In this respect there is a resemblance to the results on direct postural waking suggestion. In the latter case, however, this tendency was in the same direction as the permanent practice effects and may be nothing more than the transitory component of such practice (**5**). The case of indirect suggestibility, however, presents the paradox of the perseverational or immediate practice effects being in a direction *opposite* that of the permanent effects. If this apparent tendency should prove to be a genuine one, it

will assist materially in differentiating perseveration as an experimental entity distinct from the phenomenon of ordinary practice effects.

The bearing of these results, particularly those concerned with permanent practice effects, seems rather definitely to separate indirect suggestion from the family of reactions akin to hypnosis. This should greatly simplify the problem of defining suggestion as related to hypnotism—a thing greatly to be desired.

(Manuscript received June 4, 1932)

References

1. Binet, A., *La Suggestibilité*, Schleicher Frères, Paris, 1900.
2. Brown, W., Individual and sex differences in suggestibility, *Univ. of Calif. Pub. in Psychol.*, 1916, **2**, 291–430.
3. Hull, C. L., and Huse, B., Comparative suggestibility in the trance and waking states, *Amer. J. Psychol.*, 1930, **52**, 279–286.
4. Krueger, R. G., The influence of repetition and disuse upon rate of hypnotization, *J. Exper. Psychol.*, 1931, **14**, 260–269.
5. Patten, E. F., Switzer, S. A., and Hull, C. L., Habituation, retention, and perseveration characteristics of direct waking suggestion, *J. Exper. Psychol.*, 1932, **15**, 539–549.
6. Thorndike, E. L., *Theory of mental and social measurements*, Teachers College, Columbia University, N. Y., 1913.
7. Williams, G. W., Suggestibility in the normal and hypnotic states, *Arch. of Psychol.*, 1930, No. 122.
8. Young, P. C., Suggestion as indirection, *J. of Abn. and Soc. Psychol.*, 1931, **26**, 69–90.

FACTORS DETERMINING THE RELATIVE EFFICACY OF THE WHOLE AND PART METHODS OF LEARNING

BY ALONZO J. DAVIS AND MAX MEENES
Howard University

PROBLEM

Investigations of the whole and part methods of learning have yielded conflicting results. Some have shown the whole method superior, others the part method.[1]

With the emphasis on statistical reliability of results, conclusions on the basis of results of investigations which have shown the pure part method superior to the whole and progressive part methods have been thrown open to question. With the application of criteria of statistical reliability, all experiments in the field of verbal learning which are statistically reliable have shown that the whole method and modifications of the pure part method are superior to the pure part method.[2]

Investigators have not satisfied themselves simply with description of superiority. They have sought to find out what factors are responsible for this superiority both by hypothesis and experimentation.

Pechstein subjected to experiment several hypotheses in an endeavor to explain why the whole method and modifications of the pure part method prove superior to the pure part method.[3] His results were negative. Hence he rejected the hypotheses which he had tested and formulated a new hypothesis: namely, that all the waste in pure part learning occurs in the act of connection. This, he says, is traceable almost entirely to the influence of place association.

[1] McGeoch, G. O., *Whole-part problem:* General review and summary, *Psychol. Bull.*, 1931, 28, No. 10.

[2] McGeoch, G. O., *ibid.*

[3] Pechstein, L. A., *Whole vs. part methods in motor learning, Psychol. Monog.*, 1917, 23, No. 99.

In defining place association, he says, "Place association refers to the definite location of an element of a problem in reference not only to the remaining details of that problem, but to the entire environment. In the case of rote learning a certain syllable is learned in reference to its antecedent and consequent (immediate association) and to the remainder of the terms (mediate association). It is hereby assigned a definite position in the word series. This places it in a conceptual scheme. It is located at a definite number of syllable intervals from the introductory term and the terminal one of the list. It is reached after the same time expenditure in each trial and is followed by a constant time span for the completion of the presentation. Both spatial and temporal factors are concerned in establishing the positional relationships." [4]

In pure part learning the factor of place association does not operate so effectively as in whole learning or in modifications of the pure part method of learning. The superiority of the whole method and modifications of the pure part method Pechstein attributes to this difference in effectiveness of place association.

Crafts has more recently attempted to test the hypothesis.[5] In his investigation of card sorting in which non-serial reactions are involved, hence in which there is no place association, or only a negligible amount of it, he found the whole method, and modifications of the pure part method to be superior to the pure part method. Where place association does not operate the pure part method—it would seem—should at least prove of equal efficacy if Pechstein's hypothesis is correct.

Brown, working in the field of verbal learning, found the whole method to be superior to the pure part method.[6] He used the paired associates method in which there were 'tied up' two-syllable nonsense words with English words. His test series consisted of the elements in a 'deranged' order. It would seem that if place association operated in the learning of the material by the whole method, the whole method would

[4] Pechstein, L. A., *op. cit.*, p. 29.

[5] Crafts, L. W., Whole and part methods with non-serial reactions, *Am. J. Psychol.*, 1929, 41, 543–563.

[6] Brown, W., Whole and part methods in learning, *J. Educ. Psychol.*, 1924, 15, 229–233.

have proved at a disadvantage where the place associations were not operative.

In the two experiments cited, the emphasis has been placed on the reproduction of the material to be learned, or more specifically, on the test of reproduction. Would the same results obtain if the materials to be learned by the whole and pure part methods were equated with regard to the factor of place association? We shall attempt to answer this question by experimentation.

Other factors have been mentioned as probably determining the relatively greater efficacy of one method than the other.[7] Hunter mentions the character of the stimulus material, individual differences, the subject's age, the amount of training which he has had, and some factors as yet undetermined, as operative.[8]

With reference to the factor of individual differences, Reed, on the basis of his investigation in which the part method proved superior to the whole method came to the conclusion: "The most economical method of learning is in the last analysis a question for the individual learner." [9]

Crafts mentioned a list similiar to Hunter's.[10] In addition, he names the subject's habitual method and the subject's learning capacity. McGeoch has recently reported an investigation, in which she compared the abilities of gifted and normal children to learn by the whole, progressive part and pure part methods. She concluded: "The I. Q. is a factor conditioning the relative efficiency of the whole, progressive part and pure part methods in learning. Its efficacy as a factor is apparent, at least, in a comparison of gifted and normal children." [11]

McGeoch has recently made a critical review of all data on the problem.[12] She listed the following factors:

[7] Hunter, W. S., Experimental studies of learning, in Murchison, C., ed.: *Foundations of Experimental Psychol.*, 1929, 564–627.

[8] Hunter, W. S., *op. cit.*

[9] Reed, H. B., Part and whole methods of learning, *J. Educ. Psychol.*, 1917, **2**, 315–346.

[10] Crafts, L. W., *op. cit.*

[11] McGeoch, G. O., The intelligence quotient as a factor in the whole-part problem, *J. Exper. Psychol.*, 1931, **14**, 358.

[12] McGeoch, G. O., Whole-part problem: General review and summary, *Psychol. Bull.*, 1931, **28**, No. 10.

1. Subjects: age, training, memorizing ability and intelligence.
2. Material: type, nature, difficulty and length.
3. Practice: amount, distribution and nature of practice periods.
4. Form of practice periods used.
5. Method of measuring learning efficiency.
6. Method of measuring retention efficiency.
7. Length of interval.

Reed, who found the part method superior to the whole method, suggested that the part method adjusted the material to the memory span of the subject.[13]

The present experiment is limited to a study of the following factors: place association, sex, age, capacity and habitual method of learning.

Subjects

The subjects were 104 college students enrolled in courses in psychology. The subjects were divided into two groups to be known as Group A which consisted of 54 students, and Group B which consisted of 50 students.

Apparatus and Material

(*a*) The Malmberg Exposure Apparatus was used for the presentation of the material. This apparatus permitted a uniform rate of presentation of the material and also an equal total time for the material presented by the whole and part methods.

(*b*) The material consisted of 264 stimulus cards which were printed uniformly and legibly. The stimulus cards were 10 × 7 inches. The letters were ¾ inch × 1 inch. Upon each of 240 of the stimulus cards was printed a simple di-syllabic English word and a nonsense syllable. The nonsense syllable was printed above the English word in each case. Upon each of the remaining 24 stimulus cards was printed one of each of the di-syllabic English words referred to above.

The di-syllabic English words were chosen so as to form two sentences. The words used in Series I (referred to below) when arranged in the serial order of a sentence shall be referred to as Sentence I. The words used in Series III (referred to below) when arranged in the serial order of a sentence shall be referred to as Sentence II.

The stimulus cards were arranged in 8 series. Series I consisted of 12 sets of stimulus cards, each set of which contained 10 stimulus cards. Each stimulus card of a given set bore the same English word and nonsense syllable. Series II was composed, like Series I, except that a different group of English words and nonsense syllables was used. Series III consisted of 1 card from each of Series I arranged in the serial order. Series IV consisted of 1 card from each set of Series II arranged in the serial order of Sentence II. Series V consisted of 1 card from each set of Series I arranged in a non-serial order determined by chance. Series VI composed of 1 card from each set of Series II arranged in a non-serial order determined by chance. Series

[13] Reed, H. B., Repetition of Ebert and Meumann's practice experiment on memory, *J. Exper. Psychol.*, 1924, 15, 107–115.

VII consisted of the English words only of Series I arranged in the serial order of Sentence I. Series VIII consisted of the English words of Series II arranged in the serial order of Sentence II.

Sentence I and its nonsense syllable associates:

NIF	TOH	VED	WEC	BAJ	HEG
DAWSON	COUNTY	BANKERS	INFORMED	OVER	NINETY
FEH	KAG	LAR	MOF	BOC	VUT
COLLEGE	STUDENTS	ABOUT	KEEPING	DAILY	BUDGETS

Sentence II and its nonsense syllable associates:

HAJ	FUT	DOP	LUM	REB	KOL
CLEVER	SAXON	OUTLAWS	DECEIVED	MANY	SECRET
CEK	NUG	FOM	LUD	KEV	MIB
POLICE	AFTER	HAVING	STOLEN	SIXTY	HORSES

(*c*) The Otis Self-Administering Test of Mental Ability was given to all subjects.

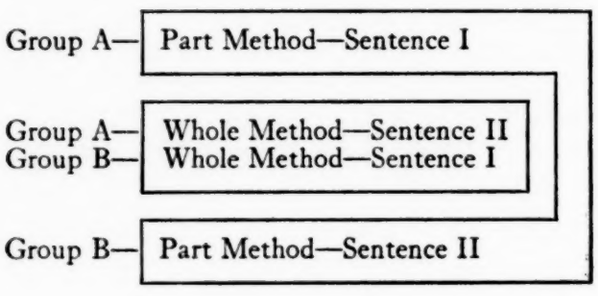
(*d*) Test booklets were constructed of fifteen numbered sheets of paper for attempts at reproduction of the nonsense syllables.

Procedure

A week before the experiment proper, the Otis Self-Administering Test of Mental Ability was given to the subjects. After the mental test they were asked to indicate their habitual method of learning by writing either the word 'part' or 'whole.' This information was obtained only after a definition by example of the whole and part methods of learning had been given. Record was not kept of those who used neither method alone. During the same period the subjects were made familiar with the paired associates method of presentation by exposing on the Malmberg Exposure Apparatus some standard cards (cards which accompany the apparatus, obtained from C. H. Stoelting Co.). Tests of reproduction were made by covering up with a piece of cardboard one of the associates on each stimulus card and by having the subjects reproduce them when their paired associates were exposed.

A week later Group A was presented with Series I by the part method, the English words being in the serial order of Sentence I. Group B was presented with Series III by the whole method. A week after this, Group A was presented Series IV by the whole method; Group B was presented with Series II by the part method, the English words being in the serial order of Sentence II. By this arrangement the benefits accruing to both the whole and part methods after practice were equalized.

The data from the two sentences presented by the whole method were pooled; the data from the same two sentences by the part method were pooled. By this arrangement, any difference in the difficulty of the material is assumed to affect the whole and part methods equally. The diagram shows how the data were pooled and the order of presentation.



Part Method

Series I was presented to Group A. The Malmberg Exposure Apparatus was so regulated that each stimulus card was exposed 0.5 second. The subjects were directed to learn the nonsense syllable associate of each English word in such a way that they would be able to write them down in their test booklets when they were presented with the Test Series, Series VII. When all of Series I had been presented, the Test Series was presented. The cards in this series were exposed by the experimenter who held them to the view of the subjects to a practiced count of about five seconds. The subjects were asked to reproduce the nonsense syllables on sheet one of the test booklets.

Series II was presented to Group B in the same manner as Series I was presented to Group A. The testing was likewise the same, with, of course, the appropriate Test Series, Series VIII.

Immediately upon completion of the test of reproduction, after exposure of the material by the part method, Series V or VI was presented. Series V followed Series I; Series VI followed Series II. It is to be remembered that Series V and VI consisted of cards arranged in a non-serial order determined by chance shuffling. Presentation and tests continued until fifteen attempts at reproduction had been made in all. The number of trials necessary for complete mastery of the nonsense syllables was scored and the average calculated. The second consecutive trial upon which a perfect reproduction occurred was considered as complete mastery.

The part method together with the whole method consisting of the stimulus cards arranged in a non-serial order shall hereinafter be referred to as the combination method.

Whole Method

Series III, a set of stimulus cards the English words of which composed Sentence I, was presented 10 times in succession to Group A. This was followed by the appropriate test series. The rate of exposure was the same as in the part method. The directions were also the same as in the part method.

Series IV, a set of stimulus cards the English words of which composed Sentence II, was presented to Group A in the same manner as Series III was presented to Group B. The test procedure was likewise the same.

Immediately upon completion of the test of reproduction, after exposure of the material by the whole method, the same materials to be learned were presented again and again by the whole method. It is to be remembered that these cards were presented in the serial order of the sentence. Each presentation was continued until fifteen attempts at reproduction had been made in all. The scoring and calculating were the same as above.

Protocols were obtained from about one-fifth of the subjects.

Tabulation of Results

I. The whole method proved superior to the part method on the criterion of the number of errors on the first attempt at reproduction.

The whole method also proved superior to the part method on the criterion of the number of individuals who found one method superior to the other. The number of individuals for whom the part method proved superior is 20. The num-

COMPARISON OF WHOLE AND PART METHODS

Method	Av. No. Errors	P.E. Av.	Diff. in Av.	P.E. Diff.
Part	9.77	.16	1.83	.21
Whole	7.94	.25		

ber of individuals for whom the whole method proved superior is 65. The number of individuals for whom the two methods proved equally difficult is 19.

The whole method proved superior to the combination method on the criterion of the number of trials necessary for complete mastery; but only by a small and unreliable difference.

COMPARISON OF WHOLE AND COMBINATION METHODS

Method	Av. No. Trials	P.E. Av.	Diff. in Av.	P.E. Diff.
Whole	8.24	.20	.57	.29
Combination	8.81	.22		

The whole method proved approximately equal to the combination method on the criterion of the number of individuals who completed learning by either method. By the whole method 74 individuals completed learning; by the combination method 75 individuals completed learning.

The whole method proved approximately equal to the combination method on the criterion of the number of individuals who found one method superior to the other. The number of individuals for whom the whole method proved superior is 39; the number of individuals for whom the combination method proved superior is 37. The number of individuals for whom the two methods proved equally difficult is 28.

II. Of 79 individuals who found one method superior to the other only 35 or 41 percent found that method superior which they habitually use.

III. When the subjects were divided into groups according to sex, there were 57 males and 47 females. Males and females learned approximately equally well with the part method.

COMPARISON OF SEXES WITH THE PART METHOD

Sex	Av. No. Errors	P.E. Av.	Diff. in Av.	P.E. Diff.
Female.....	10	.22	.4	.32
Male.....	9.6	.24		

The males were superior to the females with the whole method.

COMPARISON OF SEXES WITH THE WHOLE METHOD

Sex	Av. No. Errors	P.E. Av.	Diff. in Av.	P.E. Diff.
Female.....	8.77	.28	1.73	.43
Male.....	7.04	.33		

On the criterion of the number of individuals for whom one method proved superior to the other, 31 percent of the females for whom one method proved superior to the other found the part method superior; 18 percent of the males for whom one method proved superior to the other found the part method superior.

IV. A group composed of 43 individuals with the average chronological age of 26 learned approximately equally as well as a group composed of 53 individuals with the average age of 19 with the part method.

COMPARISON OF AGE GROUPS WITH THE PART METHOD

Av. Age	Av. No. Errors	P.E. Av.	Diff. in Av.	P.E. Diff.
19	9.7	.25	.1	.32
26	9.8	.20		

The younger group proved superior to the older group with the whole method, but with an unreliable difference.

COMPARISON OF AGE GROUPS WITH THE WHOLE METHOD

Av. Age	Av. No. Errors	P.E. Av.	Diff. in Av.	P.E. Diff.
19	7.36	.32	1.26	.47
26	8.62	.35		

V. A group composed of 53 individuals with the average gross score of 43 on the Otis S-A learned approximately equally as well as a group composed of 43 individuals with the average gross score of 57 on the Otis S-A with the part method.

COMPARISON OF CAPACITY GROUPS WITH THE PART METHOD

Av. Gross Score	Av. No. Errors	P.E. Av.	Diff. in Av.	P.E. Diff.
43	9.85	.21	.13	.35
57	9.72	.29		

A group with the average gross score of 43 on the Otis S-A learned approximately equally as well as a group with the average gross score of 57 on the Otis S-A with the whole method.

COMPARISON OF CAPACITY GROUPS WITH THE WHOLE METHOD

Av. Gross Score	Av. No. Errors	P.E. Av.	Diff. in Av.	P.E. Diff.
43	7.89	.33	.25	.45
57	7.64	.32		

INTERPRETATION OF RESULTS

I. The whole method is superior to the part method even where place association is eliminated or negligible. Several bits of evidence may be offered to show that place association was eliminated or negligible. Of protocols received from about one-fifth of the subjects only three indicated a consciousness of the horizontal or place association. The attention of the subjects was apparently directed so firmly to the vertical associations between the English words and their nonsense syllable associates that they were not conscious of the space time relationships of the items of a series which were presented.

The whole and combination methods proved approximately equal by three criteria: the average number of trials necessary for complete learning of the material, the number of individuals who completed learning by either method; and the number of individuals for whom one method proved superior

to the other. The slight difference in favor of the whole method on the criterion of the average number of trials necessary for the complete learning of the material might be explained by the fact that the combination method had the disadvantage of a larger average number of errors on the first test of reproduction, *e.g.* after presentation by the part method.

If place association was neither eliminated nor negligible in the whole method, certainly the whole method in which the stimulus cards were arranged in the serial order of a sentence—comprising a meaningful whole—would have proved reliably superior to the combination method in which the stimulus cards were presented in a non-serial order after the first test of reproduction.

II. Habitual method of learning is not a factor which determines the relative efficacy of the whole and part methods of learning. However, in the case of experiments in which only a few subjects are investigated this factor might be operative. This would be an error in sampling.

The results may have been determined partly by the reliability of each subject's report as to his habitual method of learning. Some subjects probably do not understand their own habitual method of learning although they can recognize the difference between the whole and part methods when explained to them.

III. A comparison of the data shows that the males are both relatively and absolutely superior to the females with the whole method. This result is surprising. The writers can offer no reason for the superiority except the possibility of an error in sampling which the measure of statistical reliability does not reveal.

IV. The younger group proved superior to the older group with the whole method, though with an unreliable difference. The difference in the age of the groups tested was probably not large enough to demonstrate clearly the operation of the factor of age. There were a few retarded individuals in the older group who may have lowered the average score for the group. That is to say, on the basis of evidence furnished by McGeoch,

the less intelligent do not succeed so well with the whole method as do the more intelligent.

V. The results of this investigation with regard to I. Q. as a factor in the whole-part problem confirm those of McGeoch. Though a significant difference is not found between the capacity groups with the whole method, the superiority of the whole method over the part method for the more intelligent group is relatively greater than the superiority of the whole method over the part method for the less intelligent group.

As in the case of the factor of age, the difference between the groups is small. Nevertheless a tendency in favor of the more intelligent group is shown which might be increased with a greater discrepancy in I. Q.

Conclusions

I. Place association is not a factor determining the relatively greater efficacy of the whole method over the part method. This conclusion accords with the observations of Crafts and Brown.

One of the hypotheses which Pechstein tested and rejected was disintegration through time. Though he did not find this factor operative in maze learning, it might be found to operate in verbal learning. For example, in the present investigation the number of syllables which were omitted by the subjects on the first test of reproduction was greater in the part method than in the whole method.

II. Habitual method of learning is not a determining factor.

III. Age is not a determining factor where age difference is not relatively great.

IV. Sex is a factor determining the relative greater efficacy of the whole method over the part method.

V. Capacity is not a determining factor where the groups compared do not differ significantly in gross test score. The result of this investigation confirms McGeoch's results with respect to the direction of the tendency of the superior group.

Individual differences as a determining factor is not apparent. Only one-fifth of the subjects learned better by the

part method. The subsumption of the difference in efficacy of the whole and part methods of learning under individual differences is questionable.

The relative efficacy of the whole and part methods of learning is not a resultant of place association or of the variables tested in this investigation, with the possible exception of sex. Positive and dogmatic statements about the determinants of the relative efficacy of the methods cannot as yet be made.

One possible explanation not yet tested is the operation of Jost's law which states: "Of two associations which are of equal strength but of different ages, the older receives the greater intensification from a new repetition." [14] Since the time which lapses between the repetition of each item is longer in the whole method than in the part method—it would seem —the strength of association of each item learned by the whole method is more intensified by repetition than the strength of association of each item learned by the part method, if Jost's law holds for such comparatively short intervals of time.

(Manuscript received May 28, 1932)

[14] Meumann, E., *The Psychology of Learning*, D. Appleton (1913), p. 266. .

MEMORY FOR PLEASANT, UNPLEASANT, AND INDIFFERENT PAIRS OF WORDS

BY HULSEY CASON AND FRANCES CLARA LUNGREN
University of Wisconsin

In two previous experiments,[1] Cason found that pleasant, unpleasant, and indifferent pairs of words could be learned and retained with approximately equal efficiency. The earlier studies had been few in number (**1, 3, 4, 5, 6**), and the somewhat scattered results had not been conclusive. In the present experiment the problem has been carried a step further by comparing the relative efficiency of learning and retaining pleasant, unpleasant, and indifferent pairs of words arranged in various combinations. In the previous experiments (**2**), the two words in a single pair were P, U, or I; and in the present investigation the pairs were made up of words that were P-U, U-P, P-I, I-P, U-I, and I-U to the subject. The 468 words used in the present study were the same as those used in the previous experiments.[2] All of these words were among the 3,000 most commonly used words in the English language. The P, U, and I words that were used were approximately equal in length, and the three kinds of words included practically the same per cent. of nouns, verbs, and adjectives.

Procedure.—Each subject was taken individually, and the second-named author was the experimenter for each of the 50 subjects.

Each of the 468 words had been printed on a stiff ½- by 1½-inch card, and all of the words were mixed together.

1. The S was given the collection of words; and he picked up one card at a time, looked at the word for a moment, and classified it in one of three boxes as P, U, or I, if the word itself was P, U, or I to him, or if the immediate associations were P, U, or I. The S continued to classify the words as P, U, or I until the required number had been obtained of each kind.

2. Using the words that had been classified by the S, the E arranged three sets of material, and each set consisted of 24 pairs of words. Each set included four of each of the following kinds of pairs: P–U, U–P, P–I, I–P, U–I, and I–U. The pairs

[1] **2**, Experiments II and III.

[2] The complete list of words used is given in **2**, Table 12.

of words were rotated in each set so that no one of the six different kinds of pairs would have any advantage on account of position. Each of the three sets of material, of 24 pairs each, was arranged on a stiff cardboard, and a key showing which pairs were P–U, U–P, P–I, I–P, U–I, and I–U to the S could be folded out at the side of the cardboard or it could be folded out of sight.

3. The material in a single set was presented orally two times to the S, and the pairs followed each other in an irregular order at intervals of five seconds. S was instructed to distribute his study and effort equally between all of the pairs in a single set of material, and to try to connect the words in each pair in the forward direction only during the time interval intended for that pair. Later when the first word in a pair was spoken by the E the S was to try to give the second. A special effort was made to have the S distribute his study and effort equally between the six different kinds of pairs. The conditions seemed to be as equally favorable for the six different kinds of pairs as they could be made.

4. E then tested S's learning orally, and the pairs were again taken in an irregular order. When S could not give the second word in a pair, E prompted him by repeating both of the words, and on this occasion S was given an additional five seconds to try to connect the words in the pair. S was instructed to try to learn all of the words that had been missed.

5. Procedures 3 and 4 above were followed without delay for the second and third sets of material, but in some cases the lists were presented once or three times, and the interval between E's reading of the successive pairs was occasionally seven and a half or ten seconds in length. These variations adapted the procedure to the different abilities of the different S's, and in the second and third sets of material it decreased the number of S's who learned an unusually small or an unusually large number of pairs out of a possible 24. Changing the time intervals for different S's gave whatever differences there were in the efficiency of learning the six different kinds of pairs a better chance of appearing, but it did not prejudice the results in favor of any of the different kinds of pairs.

6. A copy was made of the three sets of material; and S's retention of the pairs was again tested in the same way after an interval of one, two, or three days.

A number of precautions were taken so that during the oral presentation of the words, during the test for Learning, and during the test for Retention, the E herself would not know which words had been judged P, U, and I by the S. Every possible control was used so that the conditions would be equally favorable for learning and retaining the six different kinds of pairs.

Subjects.—A total of 50 S's was used, and most of them were undergraduate university students. Their sex and age distribution was as follows:

Age	*Men*	*Women*
30–34	2	2
25–29	2	1
20–24	8	18
15–19	4	13
Total	16	34

Results.—Table I shows the relative ranks for different S's in the tests for Learning and for Retention for the six different kinds of pairs. The figure in the upper left corner of the table

shows that 10 S's out of a total of 50 learned the P-U pairs more efficiently than the five other kinds of pairs. The relative rank in learning the P-U pairs was a tie between fourth and fifth places (a rank of $4\frac{1}{2}$) for 11 S's out of 50. The average rank in the efficiency of learning and retaining the six different kinds of pairs is shown in next to the last column of the table. The combined resulting order of rank is given in the last column. The outstanding feature of the results in Table I is the small and unreliable differences in the ranks of the six different kinds of pairs. The large amount of overlapping is much more significant than the small differences between some of the average scores.

TABLE I

RELATIVE RANKS FOR DIFFERENT SUBJECTS IN LEARNING AND RETAINING P-U, U-P, P-I, I-P, U-I, AND I-U PAIRS OF WORDS

Rank		1	1½	2	2½	3	3½	4	4½	5	5½	6	Av. Rank	Rank Order
Learning $N = 50$	P-U	10	3	1	2	8	4	1	11	3	5	2	3.35	1
	U-P	6	2	3	5	10	4	3	3	4	5	5	3.49	4
	P-I	4	2	4	5	5	6	3	5	5	4	7	3.74	6
	I-P	4	4	3	7	6	8	2	3	1	4	7	3.42	3
	U-I	5	2	3	6	3	9	4	3	4	6	5	3.64	5
	I-U	7	5	5	3	6	5	3	3	2	2	9	3.36	2
Retention after one, two, or three days $N = 50$	P-U	6	4	2	4	15	2	3	4	4	2	4	3.26	2
	U-P	4	4	5	1	11	0	9	5	7	3	1	3.43	3
	P-I	6	4	2	5	5	1	9	6	6	3	3	3.49	4
	I-P	10	5	5	3	4	3	2	5	5	6	2	3.16	1
	U-I	4	0	6	5	5	4	6	2	6	5	7	3.80	5
	I-U	6	3	6	2	5	0	3	4	3	5	13	3.86	6

The combined average results of all of the S's are shown in Table II. In the test for Learning the S's learned an average of $2.33 \pm .07$ P-U pairs of words out of a possible four; and in the test for Retention they retained an average of $1.41 \pm .06$ P-U pairs of words out of a possible four. In the test for Learning and in the test for Retention, the average scores for the six different kinds of pairs of words are approximately the same. The differences between the averages are unreliable; and the most significant feature of the results is the large amount of overlapping in the individual scores. The results do not favor the superior learning or retention of any one of the

TABLE II

RELATIVE EFFICIENCY OF LEARNING AND RETAINING P–U, U–P, P–I, I–P, U–I, AND I–U PAIRS OF WORDS: COMBINED RESULTS

		P–U	U–P	P–I	I–P	U–I	I–U
Learning	N.....	146	146	146	146	146	146
	Av.....	2.33±.07	2.36±.06	2.26±.06	2.39±.07	2.29±.06	2.28±.06
	S.D.....	1.19	1.16	1.15	1.22	1.13	1.13
Retention after one, two, or three days	N.....	146	146	146	146	146	146
	Av.....	1.41±.06	1.32±.07	1.29±.05	1.42±.07	1.15±.06	1.13±.06
	S.D.....	1.15	1.27	.95	1.19	1.03	1.04

six different kinds of pairs. The most plausible conclusion seems to be that there is little, if any, difference in the relative efficiency of learning and retaining P-U, U-P, P-I, I-P, U-I, and I-U pairs of words.

(Manuscript received July 14, 1931)

References

1. Birnbaum, K., Ueber den Einfluss von Gefühlsfaktoren auf die Assoziationer, *Monatsschr. f. Psychiat. u. Neurol.*, 1912, **32**, 95–123, 194–220.
2. Cason, H., The learning and retention of pleasant and unpleasant activities, Arch. Psychol., 1932, *21*, No. 134. Pf. 96.
3. Chaney, R. M., and Lauer, A. R., The influence of affective tone on learning and retention, *J. Educ. Psychol.*, 1929, **20**, 287–290.
4. Tait, W. D., The effect of psycho-physical attitudes on memory, *J. Abnorm. Psychol.*, 1913, **8**, 10–37.
5. Thomson, R. H., An experimental study of memory as influenced by feeling tone, *J. Exper. Psychol.*, 1930, **13**, 462–468.
6. Tolman, E. C., and Johnson, I., A note on association-time and feeling, *Amer. J. Psychol.*, 1918, **29**, 187–195.

EQUAL WEIGHTS AND PSYCHOPHYSICAL JUDGMENTS

BY LEON ARONS AND FRANCIS W. IRWIN
University of Pennsylvania

INTRODUCTION

In the method of constant stimuli, judgments are made by an S on pairs of stimuli, each pair consisting of a standard and a comparison stimulus. The standard stimulus remains constant throughout the experiment, but the comparison stimuli form a graded series extending from a stimulus value less than, to a stimulus value greater than the standard, usually including one stimulus equal to the standard. The present investigation may be considered a control experiment in which, in contrast to the usual procedure, the standard and comparison stimuli remain objectively equal. As far as the authors are aware, such an experiment has never before been performed.

METHOD

The apparatus and method of lifting were as described by Fernberger.[1] Twenty-eight cylindrical brass weights of 100 g. each were used. They were checked both at the beginning and at the end of the experiment. All weights varied less than 20 mg. from 100 g. A revolving table and a chair with arm rest and screen were employed. The rhythm of lifting was maintained by a metronome set at 92, without bell signal.

The S was seated at the revolving table. The arm rest and screen were adjusted. Instructions were given as follows:

"You will be presented with a series of weights in pairs. You are to judge the second weight of each pair as heavier than, equal to, or lighter than the first, using the terms 'heavier,' 'equal,' or 'lighter' in your report. If you are not certain of any judgment, guess."

These instructions were read to the S at the beginning of each experimental session. Seven Ss served in the experiment: Dr. S. W. Fernberger (F), Dr. H. S. Oberly (Oy), Dr. C. L. Altmaier, Jr. (A), Mr. T. J. Snee (Sn), Mr. W. N. Hallett (H), Mr. W. Carlson (C), and Mr. S. Tompkins (T).[2] With the exception of T, a graduate

[1] S. W. Fernberger: On the relation of the methods of just perceptible differences and constant stimuli, *Psychol. Monog.*, 1913, 14, 4 (Whole No. 61), 6–11.

[2] The authors wish to express their indebtedness to their Ss, and particularly to Professor Fernberger for his advice and encouragement.

student, all were members of the Department of Psychology. Of the Ss, F and Oy were very experienced in weight lifting; H and T moderately experienced; while C, S, and A were totally inexperienced, having never before participated in a weight-lifting experiment. The number of judgments obtained from each S was as follows: F—500, Oy—500, A—500, Sn—500, H—2500, C—1000, T—500. Two hundred and fifty judgments were obtained in each experimental sitting, except in the case of F who gave 200 and 300 in two sittings respectively. F, and H (during his first six sittings), the first two Ss, were allowed to pause for rest whenever they felt at all fatigued. Rest periods were then standardized, one being introduced after each series of 50 judgments. The number of rest periods taken by F corresponds closely to the standard number later allowed. H took fewer than the standard number in his first six sittings.

The lifting technique was demonstrated to the naive Ss (C, S, A) at the outset. They were allowed to practice five times using the same weight. The regular instructions were then given and the experiment was begun. No preliminary trials were given to the practiced Ss. The Ss were not informed as to the special conditions or the purpose of the investigation.

The experiment was carried on at the Laboratory of Psychology at the University of Pennsylvania during the spring and fall of 1931.

Results

The frequencies of judgments given by each S in each category will be found in Tables I and II. The results for individual sittings are given separately for each S. The 250 judgments in an individual sitting will be found fractionated into five successive series of 50 judgments. The results of F are irregular in that his two sittings consisted of four and six series of 50, respectively, and have been so tabulated. The columns headed L, E, and H contain the number of judgments of 'lighter,' 'equal,' and 'heavier' for each series of 50. In the same columns opposite P are the probability values of the different categories for each sitting as a whole, determined by the sum of the judgments in each category divided by the total number of judgments in all three categories (250 except in the case of F). The mean variations of the frequencies have been calculated from all of the judgments given by each S, on the basis of the frequencies in series of 50 judgments.

Discussion

In the discussion of the results we shall concern ourselves with the following: (1) general considerations, (2) fatigue effects, (3) the effect of previous comparisons, and (4) over-estimation.

TABLE I

FREQUENCIES OF JUDGMENTS

Subject: H (2500 j's)

Sitting	L	E	H	Sitting	L	E	H
	17	5	28		18	13	19
	8	14	28		17	19	14
1	15	10	25	6	15	19	16
	11	12	27		12	12	26
	13	13	24		16	19	15
P	.256	.216	.528	P	.312	.328	.360
	16	15	19		16	15	19
	12	16	22		20	12	18
2	12	16	22	7	12	18	20
	12	8	30		8	23	19
	16	13	21		11	17	22
P	.272	.272	.456	P	.268	.340	.392
	20	12	18		15	14	21
	9	15	26		12	20	18
3	18	12	20	8	15	12	23
	13	18	19		12	15	23
	14	12	24		12	15	23
P	.296	.276	.428	P	.264	.304	.432
	6	29	15		13	18	19
	11	21	18		11	20	19
4	8	23	19	9	13	10	27
	11	24	15		19	13	18
	8	27	15		15	10	25
P	.176	.496	.328	P	.284	.284	.432
	7	26	17		19	18	13
	13	16	21		14	19	17
5	9	16	25	10	10	13	27
	10	18	22		19	5	26
	13	19	18		11	10	29
P	.208	.380	.412	P	.292	.260	.448
				Total freq.	657	789	1054
				Mean var.	2.84	4.01	3.61
				P for total	.263	.316	.422

1. *General Considerations*

An examination of the probabilities for the various judgment categories (Tables I and II) reveals great individual differences, despite the fact that all of the weights were objectively equal. This demonstrates clearly that objective differences in stimuli are by no means necessary for difference judgments. In no case is the probability value of 'equals' (for the totals) greater than that of the other two categories, although in the case of A the values for 'equals' and 'heaviers'

TABLE II

FREQUENCIES OF JUDGMENTS (*Concluded*)

Sitting	C (1000 j's)			Sitting	Oy (500 j's)			Sitting	T (500 j's)			Sitting	F (500 j's)		
	L	E	H		L	E	H		L	E	H		L	E	H
	11	15	24		9	11	30		2	25	23		20	15	15
	13	17	20		11	9	30		2	24	24		20	12	18
1	15	19	16	1	6	13	33	1	4	19	27	1	18	12	20
	14	17	19		14	8	28		0	16	34		16	14	20
	13	19	18		11	14	23		0	27	23				
P	.264	.348	.388		.204	.220	.576		.032	.444	.524		.370	.265	.365
	16	18	16		7	14	29		2	22	26		19	13	18
	12	21	17		7	13	30		1	19	30		18	12	20
2	12	17	21	2	3	16	31	2	1	23	26		14	14	22
	13	18	19		7	13	30		2	20	28	2	18	13	19
	11	21	18		5	17	28		0	30	20		17	11	22
P	.256	.380	.364		.116	.292	.592		.024	.456	.520		24	10	16
													.364	.240	.396
				Total freq.	80	128	292		14	225	261		184	126	190
				Mean var.	2.60	2.08	1.76		0.10	3.30	2.92		1.88	1.20	1.80
				P for total	.160	.256	.584		.028	.450	.522		.368	.252	.380
					A (500 j's)				Sn (500 j's)						
	13	20	17		8	22	20		12	11	27				
	10	18	22		10	24	16		15	11	24				
3	13	15	22	1	10	18	22	1	17	6	27				
	14	14	22		11	20	19		11	13	26				
	15	15	20		14	17	19		17	6	27				
P	.260	.328	.412		.212	.404	.384		.288	.188	.524				
	11	17	22		9	21	20		13	7	30				
	14	18	18		10	16	24		16	6	28				
4	14	17	19	2	12	19	19	2	17	5	28				
	14	16	20		11	21	18		10	9	31				
	9	17	24		9	20	21		18	6	26				
P	.248	.340	.412		.204	.388	.408		.296	.132	.572				
Total freq.	257	349	394		104	198	198		146	80	274				
Mean var.	1.38	1.50	2.00		1.28	1.84	1.60		2.48	2.40	1.48				
P for total	.257	.349	.394		.208	.396	.396		.292	.160	.548				

are the same. In the case of individual sittings there are but three instances in which the probability value for 'equals' is the greatest: H's 4th, C's 2d, and A's 1st. The probabilities for individual sittings vary from .024 to .370 for 'lighters,' .132 to .496 for 'equals,' and .328 to .592 for 'heaviers.' Most of this variation is from S to S, the variation for individual Ss being relatively small as is readily seen from the mean variations. The relative amount of variability of the three categories of judgment for each S is given in Table III.

TABLE III

VARIABILITY OF JUDGMENT FREQUENCIES IN THE THREE CATEGORIES

S	L	E	H
Sn	17.0	30.0	5.4
C	10.7	8.6	10.2
A	12.3	9.3	8.1
F	10.2	9.5	9.5
T	7.1	14.7	11.2
H	21.7	25.4	17.1
Oy	32.5	16.3	6.0

These values have been computed by the formula: Variability $= MV \times 100/\text{Mean}$. In general, the category 'heavier' shows the least variability, and 'lighter' shows the greatest. For 4 Ss the category 'heavier' is the least variable and for none is it the most. Likewise, for 4 Ss 'lighter' is the most variable and for one S, the least. It seems significant that this one S (T) gave by far the smallest proportion of lighter judgments.

'Sophistication' in weight-lifting seems to have had no well-defined effect on the relative frequencies and the constancy of the results, as may be seen from the comparison of the results of F and Oy (the two most sophisticated Ss) with one another and with the other Ss.

There are no apparent trends attributable to the effect of practice. These trends might be expected to show themselves particularly with Ss H and C who gave the most extended series of judgments. Not only is this not the case, but the size of the mean variations indicates that such an influence, if present, was minimal. More extended experimentation using more Ss may, however, reveal effects due to practice.

As stated above, no S was given information with regard to the nature of the experiment. At the end of his second sitting, Oy, our most sophisticated S, stated his belief that all the weights were equal with the exception of two or three which were definitely heavier or lighter than the others. To F, there seemed to be more variability in the standard than in the comparison weights; at one time he declared that there was "no question as to the first weights differing." The other Ss seemed to take the experiment as a matter of course. From time to time, various of the Ss spontaneously remarked 'much heavier' or 'much lighter.' Thus it appears that many of the difference judgments were given with a high degree of subjective assurance.

2. *Fatigue*

In order to analyze the results for the possible influence of fatigue, they have been arranged as in Table IV. The data are given for each S separately and for the total. The results of F have been omitted from this table, since his sittings consisted of 200 and 300 judgments respectively and are, for that reason, not comparable with the others. The numbers 1 to 5 in the first column indicate the five divisions (series) of a sitting, in temporal order. The corresponding rows give the total number of judgments in each category for each of these five series. Thus, totaling the results of H for the first series in all of his ten sittings, we find that he gave 147 'lighter,' 165 'equal,' and 188 'heavier' judgments.

From an inspection of this table it is evident that while there are variations, there are no definite trends. This is true both for the individuals and for the totals. Since fatigue might be expected to manifest itself in some definite tendency relative to the five temporal divisions in the sitting, we conclude that its effect, if present, was negligible.

3. *The Effect of Previous Comparisons*

In dealing with the results, the questions arise: to what extent is the comparison-judgment made upon a given stimulus pair dependent merely upon the activity involved in

TABLE IV

FREQUENCIES OF JUDGMENTS SUMMARIZED FOR THE FIVE TEMPORAL DIVISIONS OF A SITTING

S:	H			A			C			T			Sn			Oy			Total—all Ss		
Series	L	E	H	L	E	H	L	E	H	L	E	H	L	E	H	L	E	H	L	E	H
1	147	165	188	17	43	40	51	70	79	4	47	49	25	18	57	16	25	59	260	368	472
2	127	172	201	20	40	40	49	74	77	3	43	54	31	17	52	18	22	60	248	368	484
3	127	149	224	22	37	41	54	68	78	5	42	53	34	11	55	9	29	64	251	336	515
4	127	148	225	22	41	37	55	65	80	2	36	62	21	22	57	21	21	58	248	333	519
5	129	155	216	23	37	40	48	72	80	0	57	43	35	12	53	16	31	51	251	364	483
Total	657	789	1054	104	198	198	257	349	394	14	225	261	146	80	274	80	128	292	1258	1769	2473

making *that* comparison? On the other hand, to what extent does it depend upon preceding comparisons in the series?[3] This analysis is particularly interesting since, with the present method, stimulus differences are not among the variables involved in determining the judgment, and hence do not require mathematical elimination. Conceivably, any comparison may be influenced by any previous comparison or combination of previous comparisons. While our method is adequate to study any of these influences, we have chosen the simplest, and presumably the most important, namely, the influence of the immediately preceding comparison.

The problem reduces to the question as to whether successive judgments in the various categories are distributed on a chance basis. Thus, for example, 'heavier' judgments are immediately followed sometimes by 'heaviers,' at other times by 'lighters' or 'equals.' Will the distribution of such subsequent judgments correspond with an estimation made on the basis of the observed relative frequencies for the respective categories?

With the three judgment categories, there are nine possible sequence-pairs: 'lighter-lighter,' 'lighter-equal,' 'lighter-heavier,' 'equal-lighter,' 'equal-equal,' 'equal-heavier,' 'heavier-lighter,' 'heavier-equal,' and 'heavier-heavier.' If, now, we consider the relative frequency of judgments in a given category as establishing an empirical probability of occurrence of judgments in that category, the probability of occurrence of any sequence-pair will be given by the product of the two

[3] For a general discussion of inter- and intra-serial effects, and numerous references, see W. D. Turner, Intra-serial effects with lifted weights, *Amer. J. Psychol.*, 1931, **43**, 1–25. These studies are essentially investigations of the effect of *stimulus differences* upon subsequent judgments. The effects, if any, of the comparisons themselves are not subjected to analysis. Turner's conclusion that the effects he found were 'not due to the nature of the previous *judgments* themselves' was warranted by the evidence that the effects increased with an increase in the stimulus differences. However, some effect of previous comparisons may well have been present in the form of a constant factor which his procedure was incapable of analyzing.

It may be noted that Martin and Müller, *Zur Analyse der Unterschiedsempfindlichkeit*, Leipzig, 1899, pp. 155–167, have employed a procedure similar to ours to the extent of introducing 5 comparison weights equal to the standard (Vexirversuche), in a series of 8 in order to investigate the influence of the comparison weight of one pair upon the judgment on the comparison weight of a succeeding pair.

relative frequencies concerned. For example, if the relative frequency for 'heaviers' for a given S is .50, and for 'lighters,' .30, we should expect 15 percent of all the sequence-pairs to be 'heavier-lighter,' and 15 percent 'lighter-heavier.'

In Tables V–IX will be found the results obtained by this analysis for each sitting of each S, and the results taken as a whole for individual Ss (under 'Total'). In the first column are listed the 9 possible judgment sequences, LL, for example, representing a 'lighter' followed by a 'lighter.' In the third column will be found the number of these sequences observed, as obtained by actual count from the data sheets. Pairs split by a rest period or by any other interruption have not been used in these computations. The figures in this column opposite 'Total' in the first column give the number of 'lighters,' 'equals,' and 'heaviers' involved in these pairs. The second column ('Calc.') gives the number of pairs of each kind expected on the basis of chance. These values have been obtained by multiplying the total number of pairs in each sitting by the various sequence-pair probabilities calculated as described above. In the fourth column will be found the standard errors (σ) of the calculated values computed from the formula $\sigma_{\text{freq.}} = \sqrt{pqN}$. The fifth column gives the differences between the observed and calculated values. Minus signs indicate that the number of cases observed was less than the number calculated. The slight discrepancies between the total numbers of pairs (given by 'Total L, E, H') for the calculated and observed values are to be accounted for in terms of the number of decimal places used in the computations. The sixth column gives the ratios of the differences to the standard errors, the signs of the differences being preserved. These ratios for the totals of each S have been collected in Table X.

It seems obvious from the magnitude of many of the critical ratios in the tables that judgments are not completely independent of the preceding comparisons in the series. This seems to hold for all subjects except T for whom these values are consistently low. There are individual differences in both the direction and the magnitude of the effect as dis-

TABLE V

Comparison of Observed and Calculated Frequencies of the Various Sequence-Pairs

Subject: H

Sitting	I					II				
	Calc.	Obs.	σ	D	D/σ	Calc.	Obs.	σ	D	D/σ
LL	16.3	13	3.91	− 3.3	− .84	18.8	17	4.16	− 1.8	− .43
LE	13.2	8	3.53	− 5.2	−1.47	18.0	16	4.09	− 2.0	− .49
LH	33.7	42	5.39	8.3	1.54	31.1	35	5.12	3.9	.76
Total L		63					68			
EL	13.2	11	3.53	− 2.2	− .62	18.0	13	4.09	− 5.0	−1.22
EE	10.7	18	3.20	7.3	2.28	17.5	21	4.04	3.5	.87
EH	27.1	22	4.91	− 5.1	−1.04	30.1	32	5.14	1.9	.37
Total E		51					66			
HL	33.7	39	5.39	5.3	.98	31.1	37	5.12	5.9	1.15
HE	27.1	26	4.91	− 1.1	− .22	30.1	30	5.14	− 0.1	− .02
HH	69.3	65	7.04	− 4.3	− .61	51.6	46	6.39	− 5.6	− .88
Total H		130					113			
Total L, E, H	244.3	244				246.3	247			

Sitting	III					IV				
	Calc.	Obs.	σ	D	D/σ	Calc.	Obs.	σ	D	D/σ
LL	21.6	26	4.44	4.4	.99	7.9	10	2.77	2.1	.76
LE	20.4	16	4.33	− 4.4	−1.02	22.0	17	4.47	− 5.0	−1.12
LH	31.0	31	5.20	0.0	0.00	14.3	17	3.67	2.7	.74
Total L		73					44			
EL	20.4	27	4.33	6.6	1.52	22.0	19	4.47	− 3.0	− .67
EE	19.2	19	4.21	− 0.2	− .05	61.3	70	6.79	8.7	1.28
EH	29.0	23	5.06	− 6.0	−1.19	39.8	34	5.78	− 5.8	−1.00
Total E		69					123			
HL	31.0	20	5.20	−11.0	−2.12	14.3	15	3.67	0.7	.19
HE	29.0	33	5.06	4.0	.79	39.8	37	5.78	− 2.8	− .48
HH	44.0	51	6.01	7.0	1.16	25.9	28	4.82	2.1	.44
Total H		104					80			
Total L, E, H	245.6	246				247.3	247			

Sitting	V					VI				
	Calc.	Obs.	σ	D	D/σ	Calc.	Obs.	σ	D	D/σ
LL	11.1	15	3.26	3.9	1.20	24.0	20	4.65	− 4.0	− .86
LE	19.8	15	4.26	− 4.8	−1.13	24.9	27	4.74	2.1	.44
LH	21.5	22	4.43	0.5	.11	28.2	30	4.99	1.8	.36
Total L		52					77			
EL	19.8	16	4.26	− 3.8	− .89	24.9	29	4.74	4.1	.86
EE	35.1	37	5.49	1.9	.35	25.9	24	4.82	− 1.9	− .39
EH	38.5	40	5.70	1.5	.26	29.1	27	5.07	− 2.1	− .41
Total E		93					80			
HL	21.5	21	4.43	− 0.5	− .11	28.2	28	4.99	− 0.2	− .04
HE	38.5	41	5.70	2.5	.44	29.1	30	5.07	0.9	.18
HH	42.2	40	5.92	− 2.2	− .37	32.6	32	5.32	− 0.6	− .11
Total H		102					90			
Total L, E, H	248.0	247				246.9	247			

TABLE V (*Concluded*)

Subject: H (*Concluded*)

Sitting	VII					VIII				
	Calc.	Obs.	σ	D	D/σ	Calc.	Obs.	σ	D	D/σ
LL	16.7	20	3.94	3.3	.84	15.7	19	3.83	3.3	.86
LE	22.3	19	4.50	− 3.3	− .73	19.1	19	4.20	− 0.1	− .02
LH	25.0	25	4.74	0.0	.00	27.2	24	4.92	− 3.2	− .65
Total L		64					62			
EL	22.3	20	4.50	− 2.3	− .51	19.1	19	4.20	− 0.1	− .02
EE	29.4	29	5.09	− 0.4	− .08	23.5	19	4.61	− 4.5	− .98
EH	33.3	36	5.37	2.7	.50	33.1	38	5.35	4.9	.92
Total E		85					76			
HL	25.0	25	4.74	0.0	.00	27.2	27	4.92	− 0.2	− .04
HE	33.3	35	5.37	1.7	.32	33.1	36	5.35	2.9	.54
HH	37.7	36	5.65	− 1.7	− .30	46.8	44	6.15	− 2.8	− .46
Total H		96					107			
Total L, E, H	245.0	245				244.8	245			

Sitting	IX					X				
	Calc.	Obs.	σ	D	D/σ	Calc.	Obs.	σ	D	D/σ
LL	20.1	14	4.29	− 6.1	−1.42	20.1	21	4.29	0.9	.21
LE	20.1	21	4.29	0.9	.21	18.4	14	4.12	− 4.4	−1.07
LH	30.1	35	5.14	4.9	.95	31.9	35	5.26	3.1	.59
Total L		70					70			
EL	20.1	25	4.29	4.9	1.14	18.4	18	4.12	− 0.4	− .10
EE	20.1	22	4.29	1.9	.44	16.7	18	3.94	1.3	.33
EH	30.1	23	5.14	− 7.1	−1.38	28.9	28	5.05	− 0.9	− .18
Total E		70					64			
HL	30.1	31	5.14	0.9	.18	31.9	32	5.26	0.1	.02
HE	30.1	26	5.14	− 4.1	− .80	28.9	31	5.05	2.1	.42
HH	45.1	48	6.06	2.9	.49	50.2	48	6.32	− 2.2	− .35
Total H		105					111			
Total L, E, H	245.9	245				245.4	245			

Sitting	Total				
	Calc.	Obs.	σ	D	D/σ
LL	168.4	175	12.51	6.6	.53
LE	203.3	172	13.65	−31.3	−2.29
LH	271.6	296	15.54	24.4	1.57
Total L		643			
EL	203.3	197	13.61	− 6.3	− .46
EE	245.6	277	14.87	31.4	2.11
EH	328.1	303	16.86	−25.1	−1.49
Total E		777			
HL	271.6	275	15.54	3.4	.22
HE	328.1	325	16.86	− 3.1	− .18
HH	438.3	438	18.98	− 0.3	− .02
Total H		1038			
Total L, E, H	2458.3	2458			

TABLE VI

Comparison of Observed and Calculated Frequencies of the Various Sequence-Pairs

Subject: C

Sitting	I					II				
	Calc.	Obs.	σ	D	D/σ	Calc.	Obs.	σ	D	D/σ
LL	17.6	16	4.05	− 1.6	− .40	15.7	18	3.83	2.3	.60
LE	22.5	23	4.52	0.5	.11	23.3	28	4.59	4.7	1.02
LH	25.5	27	4.78	1.5	.31	23.0	16	4.57	− 7.0	−1.53
Total L		66					62			
EL	22.5	27	4.52	4.5	1.00	23.3	26	4.59	2.7	.59
EE	28.9	24	5.05	− 4.9	− .97	34.5	22	5.45	−12.5	−2.29
EH	32.6	33	5.31	.4	.08	34.1	44	5.42	9.9	1.83
Total E		84					92			
HL	25.5	22	4.78	− 3.5	− .73	23.0	20	4.57	− 3.0	− .66
HE	32.6	38	5.31	5.4	1.02	34.1	41	5.42	6.9	1.27
HH	37.0	35	5.60	− 2.0	− .36	33.8	30	5.40	− 3.8	− .70
Total H		95					91			
Total L, E, H	244.7	245				245.7	245			

Sitting	III					IV				
	Calc.	Obs.	σ	D	D/σ	Calc.	Obs.	σ	D	D/σ
LL	16.7	12	3.94	− 4.7	−1.19	14.9	11	3.74	− 3.9	−1.04
LE	20.8	16	4.37	− 4.8	−1.10	20.3	20	4.31	− 0.3	− .07
LH	26.5	36	4.86	9.5	1.95	25.1	29	4.75	3.9	.82
Total L		64					60			
EL	20.8	28	4.37	7.2	1.65	20.3	25	4.31	4.7	1.09
EE	26.2	19	4.84	− 7.2	−1.49	27.6	24	4.95	− 3.6	− .73
EH	33.1	33	5.35	− 0.1	− .02	34.2	33	5.51	− 1.2	− .22
Total E		80					82			
HL	26.5	25	4.86	− 1.5	− .31	25.1	22	4.75	− 3.1	− .65
HE	33.1	43	5.35	9.9	1.85	34.2	41	5.51	6.8	1.23
HH	41.7	33	5.88	− 8.7	−1.48	42.7	39	5.94	− 3.7	− .62
Total H		101					102			
Total L, E, H	245.4	245				244.4	244			

Sitting	Total				
	Calc.	Obs.	σ	D	D/σ
LL	64.6	57	7.77	− 7.6	− .98
LE	87.1	87	8.91	− 0.1	− .01
LH	99.9	108	9.47	8.1	.86
Total L		252			
EL	87.1	106	8.91	18.9	2.12
EE	116.5	89	10.13	−27.5	−2.71
EH	134.1	143	10.8	8.9	.82
Total E		338			
HL	99.9	89	9.47	−10.9	−1.15
HE	134.1	163	10.8	28.9	2.68
HH	154.7	137	11.4	−17.7	−1.55
Total H		389			
Total L, E, H	978.0	979			

TABLE VII

COMPARISON OF OBSERVED AND CALCULATED FREQUENCIES OF THE VARIOUS SEQUENCE-PAIRS

Subject: F

Sitting	I					II				
	Calc.	Obs.	σ	D	D/σ	Calc.	Obs.	σ	D	D/σ
LL	27.3	23	4.85	− 4.3	− .89	39.7	34	5.86	− 5.7	− .97
LE	19.1	26	4.15	6.9	1.66	26.5	25	4.91	− 1.5	− .31
LH	26.5	24	4.79	− 2.5	− .52	41.7	49	5.98	7.3	1.22
Total L		73					108			
EL	19.1	19	4.15	− 0.1	− .02	26.5	30	4.91	3.5	.71
EE	13.5	9	3.54	− 4.5	−1.27	17.6	12	4.07	− 5.6	−1.38
EH	18.5	23	4.09	4.5	1.10	27.9	30	5.03	2.1	.42
Total E		51					72			
HL	26.5	29	4.79	2.5	.52	41.7	41	5.98	− 0.7	− .12
HE	18.5	17	4.09	− 1.5	− .37	27.9	36	5.03	8.1	1.61
HH	25.7	25	4.73	− 0.7	− .15	44.4	37	6.14	− 7.4	−1.21
Total H		71					114			
Total L, E, H	194.7	195				293.9	294			

Sitting	Total				
	Calc.	Obs.	σ	D	D/σ
LL	67.0	57	7.60	−10.0	−1.32
LE	45.5	51	6.46	5.5	.85
LH	68.5	73	7.67	4.5	.59
Total L		181			
EL	45.5	49	6.46	3.5	.54
EE	31.3	21	5.41	−10.3	−1.90
EH	46.5	53	6.48	6.5	1.00
Total E		123			
HL	68.5	70	7.67	1.5	.20
HE	46.5	53	6.48	6.5	1.00
HH	69.9	62	7.74	− 7.9	−1.02
Total H		185			
Total L, E, H	489.2	489			

TABLE VII (*Concluded*)

Subject: A

Sitting	I					II				
	Calc.	Obs.	σ	D	D/σ	Calc.	Obs.	σ	D	D/σ
LL	11.5	9	3.31	− 2.5	− .76	10.5	6	3.18	− 4.5	−1.42
LE	21.1	24	4.39	2.9	.66	19.8	19	4.27	− 0.8	− .19
LH	20.3	20	4.32	− 0.3	− .07	20.6	26	4.37	5.4	1.24
Total L		53					51			
EL	21.1	23	4.39	1.9	.43	19.8	21	4.27	1.2	.28
EE	39.2	35	5.74	− 4.2	− .73	37.0	34	5.60	− 3.0	− .54
EH	37.7	40	5.65	2.3	.41	38.5	40	5.69	1.5	.26
Total E		98					95			
HL	20.3	20	4.32	− 0.3	− .07	20.6	24	4.37	3.4	.78
HE	37.7	42	5.65	4.3	.76	38.5	42	5.69	3.5	.62
HH	36.0	32	5.54	− 4.0	− .72	39.9	33	5.78	− 6.9	−1.19
Total H		94					99			
Total L, E, H	244.9	245				245.2	245			

Sitting	Total				
	Calc.	Obs.	σ	D	D/σ
LL	22.1	15	4.59	− 7.1	−1.55
LE	41.2	43	6.14	1.8	.29
LH	41.2	46	6.14	4.8	.78
Total L		104			
EL	41.2	44	6.14	2.8	.46
EE	76.0	68	8.01	− 8.0	−1.00
EH	76.0	81	8.01	5.0	.62
Total E		193			
HL	41.2	44	6.14	2.8	.46
HE	76.0	85	8.01	9.0	1.12
HH	76.0	64	8.01	−12.0	−1.50
Total H		193			
Total L, E, H	490.9	490			

TABLE VIII

COMPARISON OF OBSERVED AND CALCULATED FREQUENCIES OF THE VARIOUS SEQUENCE-PAIRS

Subject: Oy

Sitting	I					II				
	Calc.	Obs.	σ	D	D/σ	Calc.	Obs.	σ	D	D/σ
LL	10.3	4	3.14	− 6.3	−2.01	3.4	1	1.84	− 2.4	−1.30
LE	10.8	5	3.21	− 5.8	−1.81	8.3	4	2.84	− 4.3	−1.51
LH	28.9	41	5.05	12.1	2.40	17.2	24	3.99	6.8	1.70
Total L		50					29			
EL	10.8	14	3.21	3.2	1.00	8.3	11	2.84	2.7	.95
EE	11.5	12	3.31	0.5	.15	20.6	7	4.34	−13.6	−3.13
EH	30.6	27	5.18	− 3.6	− .69	42.1	53	5.87	10.9	1.86
Total E		53					71			
HL	28.9	33	5.05	4.1	.81	17.2	17	3.99	− 0.2	− .05
HE	30.6	37	5.18	6.4	1.24	42.1	62	5.87	19.9	3.39
HH	82.3	72	7.39	−10.3	−1.39	85.8	66	7.47	−19.8	−2.65
Total H		142					145			
Total L, E, H	244.7	245				245.0	245			

Sitting	Total				
	Calc.	Obs.	σ	D	D/σ
LL	12.7	5	2.40	− 7.7	−3.21
LE	20.1	9	4.01	−11.1	−2.77
LH	46.1	65	5.65	18.9	3.35
Total L		79			
EL	20.1	25	4.01	4.9	1.22
EE	31.4	19	6.14	−12.4	−2.02
EH	72.5	80	8.35	7.5	.90
Total E		124			
HL	46.1	50	5.65	3.9	.69
HE	72.5	99	8.35	26.5	3.17
HH	168.1	138	10.50	−30.1	−2.87
Total H		287			
Total L, E, H	489.6	490			

TABLE VIII (*Concluded*)

Subject: T

Sitting	I					II				
	Calc.	Obs.	σ	D	D/σ	Calc.	Obs.	σ	D	D/σ
LL	0.2	1	0.49	0.8	1.62	0.0	0	0.00	0.0	—
LE	3.7	3	1.90	− 0.7	− .37	2.2	1	1.48	− 1.2	− .81
LH	4.2	4	2.03	− 0.2	− .10	2.5	4	1.56	1.5	.96
Total L		8					5			
EL	3.7	3	1.90	− 0.7	− .37	2.2	4	1.48	1.8	1.22
EE	47.5	49	6.19	1.5	.24	51.2	50	6.38	− 1.2	− .19
EH	56.8	56	6.61	− 0.8	− .12	58.6	58	6.68	− 0.6	− .09
Total E		108					112			
HL	4.2	4	2.03	− 0.2	− .10	2.5	2	1.56	− 0.5	− .32
HE	56.8	58	6.61	1.2	.18	58.6	60	6.68	1.4	.21
HH	68.1	67	7.01	− 1.1	− .16	66.6	66	6.96	− 0.6	− .09
Total H		129					128			
Total L, E, H	245.2	245				244.4	245			

Sitting	Total				
	Calc.	Obs.	σ	D	D/σ
LL	0.5	1	2.21	0.5	.23
LE	5.9	4	2.41	−1.9	−.79
LH	6.9	8	2.60	1.1	.42
Total L		13			
EL	5.9	7	2.41	1.1	.46
EE	99.0	99	8.89	0.0	.00
EH	115.2	114	9.39	−1.2	−.13
Total E		220			
HL	6.9	6	2.60	−0.9	−.35
HE	115.2	118	9.39	2.8	.30
HH	134.8	133	9.88	−1.8	−.18
Total H		257			
Total L, E, H	490.3	490			

TABLE IX

COMPARISON OF OBSERVED AND CALCULATED FREQUENCIES OF VARIOUS SEQUENCE-PAIRS

Subject: Sn

Sitting	I					II				
	Calc.	Obs.	σ	D	D/σ	Calc.	Obs.	σ	D	D/σ
LL	20.6	19	4.35	− 1.6	− .37	22.0	15	4.47	− 7.0	−1.57
LE	12.7	8	3.48	− 4.7	−1.35	9.7	7	3.05	− 2.7	− .89
LH	37.5	44	5.64	6.5	1.15	41.4	51	5.86	9.6	1.64
Total L		71					73			
EL	12.7	13	3.48	0.3	.09	9.7	11	3.05	1.3	.43
EE	7.8	6	2.76	− 1.8	− .65	4.1	2	2.01	− 2.1	−1.04
EH	23.5	25	4.61	1.5	.33	18.2	19	4.10	0.8	.20
Total E		44					32			
HL	37.5	39	5.64	1.5	.27	41.4	46	5.86	4.6	.78
HE	23.5	31	4.61	7.5	1.63	18.2	23	4.10	4.8	1.17
HH	69.1	60	7.04	− 9.1	−1.29	77.4	68	7.26	− 9.4	−1.29
Total H		130					137			
Total L, E, H	244.9	245				242.1	242			

Sitting	Total				
	Calc.	Obs.	σ	D	D/σ
LL	42.9	34	6.25	− 8.9	−1.42
LE	22.4	15	4.62	− 7.4	−1.60
LH	78.9	95	8.13	16.1	1.98
Total L		144			
EL	22.4	24	4.62	1.6	.35
EE	11.7	8	3.38	− 3.7	−1.09
EH	41.4	44	6.15	2.6	.42
Total E		76			
HL	78.9	85	8.13	6.1	.75
HE	41.4	54	6.15	12.6	2.05
HH	146.1	128	10.11	−18.1	−1.79
Total H		267			
Total L, E, H	486.1	487			

TABLE X

D/σ FOR TOTALS FOR EACH S

S	H	C	F	A	Oy	T	Sn
LL.....	.53	− .98	−1.32	−1.55	−3.21	.23	−1.42
LE.....	−2.29	− .01	.85	.29	−2.77	−.79	−1.60
LH.....	1.57	.86	.59	.78	3.35	.42	1.98
EL.....	− .46	2.12	.54	.46	1.22	.46	.35
EE.....	2.11	−2.71	−1.90	−1.00	−2.02	.00	−1.09
EH.....	−1.49	.82	1.00	.62	.90	−.13	.42
HL.....	.22	−1.15	.20	.46	.69	−.35	.75
HE.....	− .18	2.68	1.00	1.12	3.17	.30	2.05
HH.....	− .02	−1.55	−1.02	−1.50	−2.87	−.18	−1.79

closed by these results. The only pairs of judgments showing consistency of direction are the LH's which are positive, and the HH's which are negative, for all Ss. In addition there seems to be a tendency in 5 Ss (C, F, A, Oy, Sn) toward avoiding the making of two consecutive judgments in the same category. For these Ss the values for LL, EE, and HH are not only negative in all cases but in general are fairly large. The consistency of direction of these values, we believe, gives additional evidence for their reliability beyond that indicated in the ratios themselves. On the basis of our data, which may be regarded chiefly as illustrative of the possibilities of our technique, we are unable to generalize further. It remains for future experiment to determine the true extent both of individual differences and of common trends.

4. *Overestimation*

The results indicate the operation of another factor, that of overestimation of the second weight. Referring to Tables I and II, we find that in the case of each S the total frequency of 'heaviers' exceeds that of 'lighters.' Except for F, this difference is large for all Ss. This is much as would be expected in view of the well-known fact of overestimation of the second stimulus in psychophysical procedures. The problem arises as to the relationship between this overestimation and what generally is called time error. This problem is somewhat beyond the scope of the present paper, and requires special investigation. This is now in progress.

Summary and Conclusions

Psychophysical judgments on objectively equal weights were obtained from seven Ss. On the whole, 'heavier' judgments showed the least variability and 'lighter' judgments the most. An analysis of the results revealed no apparent effects attributable to 'degree of sophistication' of the Ss, practice, or fatigue. The effect of previous comparisons, however, seemed quite definite in many cases. Five Ss, for example, showed a tendency toward avoidance of the repetition of judgments in the same category. However, the necessity for further research was indicated. All Ss showed evidence of overestimation of the second weight.

Our technique is offered as one particularly well adapted to the study of certain factors involved in the making of psychophysical judgments, such as the effect of previous comparisons, which are difficult of analysis by the usual methods. It is suggested that the method may prove fruitful in attacking factors which have not been discussed in this paper, such as attitude, the effect of instructions, and temperamental type, as well as the investigation of other variables which can be attacked best by means of simplified experimental conditions such as are embodied in this technique, rather than by the more complex settings of the customary psychophysical procedures.

(Manuscript received October 8, 1931)

INFLUENCE OF AN INTERPOLATED ELECTRIC SHOCK UPON RECALL

BY M. M. WHITE
University of Kentucky

The effect of interpolated emotional stimuli upon recall has been studied by Robinson, Skaggs, Harden, Stone, and others. All agree that the interpolation affects recall; but the nature of the effect is not clear. Harden (**5**) found that an electrical stimulus aided the recall of eleven subjects, hindered it in twenty-five, and had no effect on eight. What was the cause of the varying effect? A possibility is that it is due to varying degrees of the intensity of the emotional disturbance.

One measure of the intensity of emotional disturbance that has been used in the past is the galvanic reflex. To be sure there is no agreement among psychologists as to its validity. There is agreement, however, that the galvanic reflex occurs during emotional disturbances. The present writer sought by the use of the galvanometer to discover whether the effect upon recall of the interpolated emotional situation is related to the amount of disturbance as measured by the galvanometer.

Procedure

Four lists of fifteen nouns of three letters each were used. Each list was presented one time to each subject on the Ranschburg memory apparatus. Each word was exposed to view for two seconds with no appreciable intervening interval. After an exposure the subject was either given an electrical stimulus and then read a book for twelve minutes or read immediately for twelve minutes. As the shock delayed the reading no more than ten seconds no substitute activity other than reading was provided.

The subject was requested not to think about the words during the time he read the book. After the twelve minute interval he was given pencil and paper and asked to write down all the words he remembered in the list. When the subject had finished he indicated the fact and the time he had spent in recall was recorded. The subjects were allowed to work as long as they wished, and were not aware that they were being timed.

After he had written all the words he could, the subject was told that he would not be tested again on that particular list, and instructed to forget the words as much as possible. He then read for five minutes.

At the end of five minutes a new list was presented and the above procedure was repeated until each subject had seen four lists after two of which he was shocked. The order of presentation of the lists and the position of the shock varied from subject to subject. For example, if one subject was given the lists in the order of D C A B and shocked after D and after A the next subject might be given the lists in the order of A B C D and shocked after B and C. The lists were given to every subject in a different order. By this procedure we attempted to control such factors as practice, the relative difficulty of the lists, fatigue and any possible effect of the shock due to position in the series. Those lists after which the shock was given are referred to as the experimental lists; the others are the control lists. In treating the results of any given subject the experimental lists are combined and the control lists are combined. A few subjects were given only two lists, receiving a shock after one of them.

The galvanometer was connected as follows: The electrodes were made of zinc and were about three centimeters in diameter and two millimeters in thickness. These electrodes were covered with a paste made of kaolin, an extremely weak solution of zinc sulphate, and water. The hook-up used was one suggested by Darrow (2, 418–420). Briefly it is as follows: The leads from the electrodes were connected in series with a ten thousand calibrated ohm resistance and in turn to the X arm of the wheatstone bridge. A shunt was inserted across the galvanometer poles. A one hundred thousand ohm resistance was so hooked up that it could be substituted for the subject; that is, a two-way switch was inserted so that either the subject and the ten thousand calibrated resistance were in the X arm of the wheatstone bridge, or the one hundred thousand ohm resistance and the calibrated resistance of ten thousand ohms were in the X arm of the wheatstone bridge. The deflections of the galvanic mirror were traced on a moving piece of paper. The paper was sixteen centimeters wide and was ruled lengthwise—a line occurring every two centimeters.

The procedure for measuring the galvanic resistance was as follows: The electrodes were attached; the wheatstone bridge was balanced with the resistance in series with the subject at five thousand ohms. The subject was allowed to rest ten minutes. The wheatstone bridge was again balanced. The one hundred thousand ohm resistance was substituted for the subject and adjusted until it balanced the wheatstone bridge. The shunt of the galvanometer was adjusted until a two centimeter deflection was caused by a change of five thousand ohms. Then the subject was substituted for the one hundred thousand ohm resistance. The words were presented, and the galvanic deflections were followed on the moving paper. In the experimental series the shock was given immediately after the list of words was presented. The instant it was given was marked on the moving paper and the galvanic deflections recorded for ten seconds after the shock. The one hundred thousand ohm resistance was substituted for the subject once more and changed until the wheatstone bridge was balanced. The calibrated resistance was then changed until the maximum deflection of the galvanic mirror as traced on the paper was reproduced. This measurement represented the amount the subject's resistance changed after the shock. In order to make subjects comparable the ratio of this change to the subject's resistance just before the words were presented was computed and stated in terms of percent.

Thirty subjects were used; twenty-two men and eight women. The subjects were students in the elementary psychology classes.

Results

From Table I we see that: (1) thirteen subjects recall more words in the experimental series than in the control

TABLE I

VARIATIONS IN NUMBER OF WORDS RECALLED, TIME, AND GALVANIC RESISTANCE

	All subjects		S's who improved		S's who did worse		S's who did the same	
	Exper.	Control	Exper.	Control	Exper.	Control	Exper.	Control
No. of subjects	30	30	13	13	15	15	2	2
Av. number of words recalled	6.93	6.33	9.54	5.23	4.73	7.27	6.5	6.5
Av. time per word recalled (seconds)	21.59	15.98	13.97	19.16	22.70	20.94	18.85	10.38
Av. percent change G. R.	41.16		56.90		25.26		37.00	

series; fifteen recall fewer words in the experimental than in the control series; and two subjects recall the same number of words in the two series. (2) The average gain shown by the subjects who improve is greater than the loss shown by the subjects who are hindered by the shock. This is true whether we compare the experimental series with the control series within the same group or with the total number of subjects. (3) The average time per word recalled is reduced after shock for the subjects who gain in recall but is greater for subjects who lose or remain the same. (4) The average change in electrical resistance in terms of percent is twice as great for the subjects who gain in recall after shock as for the subjects who lose (56.90 as compared to 25.26).

The relation between change in resistance and recall after shock is shown in Chart I. Thirty-six percent may be looked upon as a rough critical score. Of the sixteen subjects whose change was less than thirty-six percent, twelve showed a loss

CHART I

ELECTRICAL RESISTANCE AND RECALL

Change of galvanic resistance	Loss in words	Gain in words
100% – 62%	0 S's	7 S's
37%	3 S's	3 S's
0%	12 S's	3 S's

in recall, three showed a gain, and one remained constant. Of the thirteen subjects whose change was greater than thirty-six percent, ten showed a gain in recall after shock, three showed a loss and one remained constant. The seven subjects whose resistance changes were greatest also showed the most improvement in recall after the shock; that is, the seven subjects whose resistance changed as much as sixty-two percent did better during the experimental series than in the control series.

The explanation of the varying effect of the electrical stimulus may lie in the following factors: 1. The factors of retroactive inhibition, 2. Changes in tonicity, 3. Absence or presence of an emotion, 4. Unitary experience of the electrical stimulus and the words to be learned.

In the case of the fifteen subjects who were hindered by the electrical stimulus, the explanation must lie in those factors which explain retroactive inhibition, such as breaking of set, as indicated by Mueller and Pilzecker. These factors of course operated also in the case of the thirteen subjects who improved as the result of the electrical stimulus. In these subjects it is to be remembered the electrical stimulus elicited the greatest galvanic change. They probably had the greatest increase in tonicity as the result of the electrical stimulus. The evidence for the higher tonicity is of two kinds: first, lowered resistance accompanies or is the result of increased muscular tension (**8, 9**); second, the speed of work was greater in these subjects. The electrical stimulus and subsequent higher tonicity occurred so closely after the presentation of the series of words that they were experienced together.

The above explanation is in accord with the work of Bills (**1**) who found that more efficient work was done by subjects when they were tense than when they were relaxed. If our supposition is correct Bills would have had similar findings if his subjects had increased their tenseness immediately after the presentation of the words, though the effect would not have been so pronounced.

It has been suggested by some experimenters (**6**) that whereas large changes in electrical resistance indicate the

presence of emotion small changes do not. Darrow (3) suggests that emotions may have two aspects, one "the immediate reflex response to 'sensory' excitement" and the other "the response mediated by associative processes or ideas aroused by the stimulus." It is conceivable that if only the 'immediate reflex response' occurs there is only a small galvanic change and that there is no emotion. In our subjects those whose galvanic change was less than thirty-six percent experienced no emotion; those whose galvanic change was relatively great did experience an emotion. If this view is correct, our results are in accord with the suggestion given by Dashiell (4) that mild emotions are helpful though extreme emotions are a hindrance to efficiency.

In the case of the two subjects who were neither helped nor hindered by the electrical stimulus, there must have been a balancing of facilitating factors and retroactive inhibitory factors.

Summary

Electrical stimuli were interpolated between learning and recall. The galvanic deflection was recorded. It was found that:

1. Electrical stimuli occurring within ten seconds after the presentation of a list of words to be learned facilitated recall in thirteen out of thirty subjects; hindered recall in fifteen subjects; and had no effect upon two subjects.

2. Relatively small changes in galvanic resistance accompanied a loss in recall and relatively large changes in galvanic resistance accompanied a gain in recall.

3. The average time per word for recall after shock was less for subjects who showed a gain in the number of words than in any other series.

An explanation in terms of increased tonicity is suggested.

(Manuscript received July 25, 1931)

References

1. Bills, A. G., Influence of muscular tension on the efficiency of mental work, *Amer. J. Psychol.*, 1927, **38**, 227–251.

2. Darrow, C. W., Continuous quantitative records in standard units by the use of the wheatstone bridge method, *J. Gen. Psychol.*, 1930, **4**, 418–420.

3. Darrow, C. W., Differences in the physiological reactions to sensory and ideational stimuli, *Psychol. Bull.*, 1929, **26**, 185.
4. Dashiell, J. F. *Fundamentals of Objective Psychology*, 1928. Houghton Mifflin. Page 362.
5. Harden, L. M., The effect of emotional reactions upon retention, *J. Gen. Psychol.*, 1930, **3**, 1927–221.
6. Landis, C., and DeWick, H. N., The electrical phenomena of the skin, *Psychol. Bull.*, 1929, **26**, 64–119.
7. Müller, G. E., and Pilzecker, A., Experimentelle Beitrage zur Lehre vom Gedachtnis, *Zsch. f. Psychol.*, 1900, Erg., **1**, pp. 300.
8. Richter, C. P., Significance of the changes in electrical resistance of the body during sleep, *Proc. Nat. Acad. Sci.*, 1926, **12**, 214–222.
9. White, M. M., Relation of bodily tension to electrical resistance, *J. Exper. Psychol.*, 1930, **13**, 267–277.

OCULAR DOMINANCE IN YOUNG CHILDREN

BY RUTH UPDEGRAFF

Iowa Child Welfare Research Station
State University of Iowa
Iowa City, Iowa

That most individuals are inconvenienced to such a slight degree by normal diplopia is not surprising. The fact that our attention is primarily concerned with the object upon which our eyes are focused is an important factor in making us disregard the double images of other objects in front of that object and behind it. Moreover, when forced to consider two objects in relation to each other we are able to disregard the image of one eye and attend only to the image of the other. One eye dominates the other, so to speak. Of this many of us are quite unaware.

Although this ignoring of one image in order that the other may become clarified has been recognized for some centuries, it is only in recent years that ocular dominance has been studied with any consistency or exactness.[1] Interest in ocular dominance has increased with knowledge of dominance of other types, particularly hand dominance. The possibility of the existence of relationships between various types of dominance has encouraged analytic studies in this field.

It is true that under certain conditions an individual is forced to use one eye or the other. This applies not only to such circumstances as those in which sighting is involved, but also to others in which both eyes are in use and directed at the same object. The individual is not conscious that the image of one eye dominates that of the other, for to the best of his knowledge he is seeing with both eyes. It has been fairly well established in the case of adults that one of the eyes is consistently dominant in those situations in which there must

[1] Dr. Walter Miles has contributed two discussions of the history of the study of this subject (2, 3).

be a choice between the two images. Just why this is true is still unanswered. The present study was undertaken with the hope of determining how early in life this dominance may be found.

Nearly all of the methods used in previous studies of ocular dominance are unsuitable for experimental work with young children inasmuch as they necessitate verbal responses based upon judgments, or subjective accounts of the visual field. It seemed, however, that the underlying principle of the use of the V-scope designed by Dr. Miles (**2**) was feasible. This simple apparatus is easily manipulated by the observer. It consists of a truncated, flattened cone, 12 inches long, large enough at one end to enclose the forehead and both eyes, and small enough at the other so that the visual field 6 feet away is limited to a diameter of some 12 inches. Looking through this cone at the object 6 feet distant the observer cannot see it with both eyes because the opening of the cone is too small. He must so hold the cone that one eye or the other is used. He is quite unconscious of this, however, and since he is asked to make some judgment concerning the distant object, his attention is directed at it and not at the V-scope. To the observer the V-scope is a means, not an end. On the other hand the examiner, who is holding the object, can readily see which eye the subject is using by noting which eye shows through the small end of the cone. The dominant eye is thus determined independently of verbal report. There is, also, less chance for one hand or the other to influence the result since both hands are needed to hold the V-scope.

Method of Experiment

Subjects

The children participating in this study were all in attendance in the preschool laboratories of the Iowa Child Welfare Research Station. One hundred ninety children from two to six years of age were tested. One hundred thirty-three of these were tested twice, thirty-eight three times, eleven four times, and one five times; twenty-three were given additional retests in the first and second grades.

Apparatus

At the time this study was begun, an attempt was made to use the linen V-scopes originated by Miles (**2**). It was found that although they were fairly satisfactory for the testing of five-year-old children, they were not fulfilling their purpose with the younger children. The collapsibility of the original V-scope, an advantage in the testing of older subjects, is a real problem for the young child. He finds the V-scope so difficult to keep open that he cannot look through it as he wishes. To overcome this disadvantage metal tubes were made. Sheet-zinc cones, identical to the V-scope in dimensions and pattern, proved easy for the children to hold. Edges of the metal were turned back and pressed down so that all surfaces were smooth. The metal cones had an additional advantage in that they could be disinfected easily.

The objects to be viewed were the ten small pictures forming part of the Miles test. These are outline sketches on 2 × 3 inch cards of such figures as a woman's head, a cat, a bicycle. Although all of these drawings are not particularly suited to the interests of young children, some of them are sufficiently simple and familiar to be comprehensible even to the youngest ones; others need careful scrutiny by the oldest. There is, therefore, a quite satisfactory mixture of easy material to give encouragement, and difficult material to insure careful inspection by the child. It will be remembered that for test purposes these pictures are important only instrumentally and the correct naming of them by the children has no significance.

Procedure [2]

The child is brought into the testing room where there are two small chairs facing each other, 6 feet apart. On a table nearby are three V-scopes. The examiner picks up one of these, saying, "Do you see this? It's a long thing for you to look through. We call it a V-scope. See—you hold it up like

[2] For the collection of approximately one-fourth of the data reported in this study the author is indebted to her class in experimental psychology of childhood, 1929–1930, and to Miss Lois Jack, research assistant in the Iowa Child Welfare Research Station.

this." (The examiner demonstrates, sitting down as he does so.) "You try it. Now I'm going to show you some pictures and you can see what they look like through the V-scope. Suppose you sit in this chair." The examiner sees that the child is comfortably seated in the chair directly facing him. ["Now you hold this in your hands" (placing the V-scope in his hands in the correct manner) "and see if you can find the picture I shall hold in front of my nose." (The examiner holds the picture as indicated and waits for the child to find it and look at it.) "Good. Now I'm going to give you another V-scope."] The examiner changes the V-scope. "Now, I'm going to sit over here and show you some pictures. You tell me what they are."

The examiner shows ten pictures, recording whether the child is looking with the left or the right eye. The V-scope is changed between each picture so that a new adjustment is assured each time. Before a picture is shown the examiner repeats the part of the instructions included between the brackets.

During the first two years this test was given, all the children were shown ten pictures. Since it was found that if a child shifted in the use of eyes he did so early in the series, in the third year of the experiment the plan of presentation was slightly changed; only the first five pictures were shown to those children who responded similarly in all five. If there was a shift in use of the eyes all ten pictures were shown.

If a child used the same eye eight, nine, or ten times in the ten chances he had, he was designated as using that eye dominantly. Those children who did not respond so decidedly but who used one eye nearly as frequently as the other, that is, in five, six, or seven times out of ten, were said to be indefinite in eye dominance.

Results

The experiment proved to be an interesting one to the children. The metal tubes were easy to handle, there seemed to be an added motivation in the fact that the V-scopes had to

be changed for each picture, and the task of identifying the pictures seemed to be challenging.

Reliability

In order to determine whether the results of one test might be considered representative, repeated tests were given at intervals of not more than two months. The following tabulation presents the data thus secured.

Age, Years	Children	Same Eye Dominant			Different Eye Dominant
		Left	Right	Both	
2	13	2	0	2	9
3	19	6	9	1	3
4	27	8	15	1	3
5	45	10	27	3	5
Total	104	26	51	7	20

The percentage of agreement between first and second tests of three-, four-, and five-year-old children is high. Dominance was the same in the case of 88 percent of these children. Apparently the test was not reliable for two-year-old children since so large a proportion of their number reacted differently in the two tests. This may mean that the test is not a satisfactory one for such young children or that it is suitable but dominance has not been established. In addition, the number of cases is small. In consideration of the above facts the remainder of the study is limited to the results of tests of three-, four-, and five-year-old children.

The reliability of the test, when used with the three older ages, compares well with that found by Miles (**2**) for first and third grade children (namely, from 75 to 95 percent), and for children and adults (**3**) (approximately 80 percent).

Frequency of Dominance Types

The following tabulation showing the relative frequency of occurrence of dominance types in three-, four-, and five-year-old children does not include data from tests found unreliable:

Age, Years	Children	Percent		
		Left Eye	Right Eye	Indefinite
3	39	20.5	64.1	15.3
4	71	22.5	66.1	11.2
5	88	35.2	57.9	6.8

Evidently the relative frequency of right- and left-eyedness as shown by this test does not vary significantly with age at these early years. Right-eyedness occurs some two and a half times as often as left-eyedness. Here, as well as in reliability, there is agreement with the results of other investigators observing older subjects. Parson (**4**) using his manuscope, found 69 percent of 877 public school children right-eyed, 29 percent left-eyed. By a similar method, Cuff (**1**) obtained with a group of 146 school children percentages of 62 and 29 respectively. Miles' (**2**, **3**) right-eyed subjects represented approximately 64 percent of his group, while he found approximately 34 percent left-eyed.

There seems to be some indication that indefinite eyedness or the tendency to use one eye at some times, the other eye at other times, decreases with age in the early years. This is indicated in the present study and the conclusion is strengthened by data of others. This study shows fewer instances of indefinite or impartial eyedness at five years than at three. In older subjects Parson found indefiniteness in 1 percent of his cases, Cuff in 9 percent, Miles in 2 percent.

From the above results concerning reliability and frequency of eye dominance, it seems justifiable to believe that for many children ocular dominance, as indicated by this test, is not established before the age of three, and that some children do not show dominance for several years afterwards. It is not known whether dominance is ever established in the case of some of these individuals. Pertinent results in this study are discussed later.

Sex Differences

Table 1 indicates that there are no sex differences greater than the possibility of error.

TABLE 1

OCULAR DOMINANCE IN THREE-, FOUR-, AND FIVE-YEAR-OLD CHILDREN

Age, Years	Children	Percent		
		Left Eye	Right Eye	Indefinite
Boys				
3	26	23.0	57.6	19.2
4	35	17.1	71.4	11.4
5	42	35.7	59.5	4.8
Girls				
3	13	15.3	76.9	7.7
4	36	27.8	61.1	11.1
5	46	34.8	56.5	8.6

Consistency of Eyedness Over a Period of Time

In order to throw more light on the problem of stability of eyedness, sixty-three children were given several successive tests, the total intervening time varying from six months to two years. The following tabulation shows the results:

Tests	Interval Between Tests	Children	Same Eye Dominant			Dominance Shifted
			Right	Left	Indefinite	
3	1 year.....	13	6	1	0	6
2	1 year.....	39	18	8	2	11
2	6 months.....	11	4	4	0	3
Total		63	28	13	2	20

Additional information was gained from tests given to twenty-three six- and seven-year-old children who had been last tested at the age of five. Of the twenty-three, seventeen used the same eye dominantly as in the previous tests. In the whole group there were fifteen right-eyed children, seven left-eyed, and one apparently able to use either. The same children in the last preceding test were classified as follows: 12 right-eyed, 10 left-eyed, and one who used either eye. A larger percent of the children originally left-eyed had shifted than was the case with those originally right-eyed.

Earlier in this paper the question was raised concerning the

eventual establishment of dominance in children who were impartial-eyed at the age of three. Of the sixty-three children in the above table eight were impartial-eyed at the first test. At the time of the last test, two of these still remained in that classification, while six had become either right- or left-eyed.

Relation to Handedness

Previous investigators who have studied the relation between hand and eye dominance have found that right-eyed persons are more apt to be right-handed than are the left-eyed to be left-handed. The incidence of left-eyedness is greater than that of left-handedness. For several reasons it is not possible to be definite about these dominance relationships in young children. There are as yet no proved methods of determining handedness nor exact knowledge of its nature. Handedness tests of young children have included anatomical measurements, tests of strength, skill, and customary usage. Investigators have been prone to choose a single criterion, or, in the face of varying results with several criteria, have been in doubt concerning correct interpretation.

After giving a battery of tests of handedness to seventy-five of the children used in this study, it is possible to state that such handedness tests as are available indicate results similar to those found in the case of adults. However, the reliability of the tests in the battery varied; some were not validated by observational data. At the present time a more comprehensive study of handedness is in progress and it is hoped that in the near future valid and reliable data upon which to base correlative studies will be available.

In the course of this study two cases of interest in relation to this matter were noted. Between the ages of four and seven two children shifted in hand preference—one from right to left, the other from left to right. This information is based on teacher reports. At the time their ocular dominance was first determined each was identically dominant in eye and hand. In the following year, the year of transition in handedness, the test showed no eye dominance present. The next test indicated that the ocular dominance had shifted with the hand. One of these children was the subject of his parents'

attempts to change his handedness. The circumstances of the other instance are not known.

Summary

1. A modification of the Miles A-B-C Vision Test was used in 396 tests of children two to six years of age.

2. The test proved to be reliable for children from three to six years of age.

3. This method did not reliably test the two-year-old children to whom the test was given. This may mean that the test is not a satisfactory one for such young children, or that it is suitable but dominance has not been established.

4. The relative frequency of right-eyedness and left-eyedness in children from three to six years old is comparable to the frequency of their occurrence in older children and adults, as found by previous investigators. Approximately 60 percent of the children were right-eyed and from 20 to 30 percent left-eyed.

5. Results indicated that impartial-eyedness is more common in young children than in older. It occurs more frequently in three- than in five-year-olds. From a comparison of these results with those of others studying older children, indications are that impartial-eyedness is more common in preschool children than in children of school age.

6. There were no apparent sex differences.

7. It seems justifiable to believe that for many children ocular dominance as indicated by this test is not established before the age of three. Thereafter, however, it is commonly found.

8. Consistency in ocular dominance over a period of years was apparent in approximately 75 percent of the children of whom such testing was possible.

(Manuscript received May 19, 1932)

References

1. Cuff, N. B., The interpretation of handedness, *J. Exper. Psychol.*, 1928, 11, 27–39.
2. Miles, W. R., Ocular dominance demonstrated by unconscious sighting, *J. Exper. Psychol.*, 1929, 12, 113–126.
3. Miles, W. R., Ocular dominance in human adults, *J. Gen. Psychol.*, 1930, 3, 412–430.
4. Parson, B. S., *Left handedness, a new interpretation.* New York: Macmillan, 1924. Pp. vii, 185.

THE AUTONOMIC NERVOUS SYSTEM AS A FACTOR IN THE PSYCHOGALVANIC REFLEX

BY W. D. O'LEARY, S.J., M.D.
St. Louis University

The part played by the autonomic nervous system in the production of that somewhat enigmatic phenomenon known as the psychogalvanic reflex has been frequently emphasized by many authors. The various theories held at the present time concerning this and other aspects of this reflex may be found summarized in a very satisfactory manner by Landis and DeWick.[1]

The object of this paper is not to add another theory to those already existing but to present some rather interesting laboratory findings. Experimenting on ourselves and students in the psychological laboratory of St. Louis University we focused our attention on factors which might influence the autonomic, eliminating as far as possible all emotional stimuli and muscular movements. The factors selected as most likely to produce appreciable effects on the autonomic were very deep breathing, shallow and rapid breathing, the holding of the breath and increased heart action. These have been used by other investigators, notably Radecki.[2] With this author, Vigouroux and Piéron, we also felt that vasodilation should give considerable information.

The results of these experiments were far from satisfactory for although we obtained changes in resistance, and often very marked changes, there seemed to be no way of assuring ourselves that we were measuring only autonomic changes to the exclusion of others such as the hydrogen ion concentration of the blood.

The only significant thing shown by these experiments on

[1] Landis, C., and DeWick, H. N., The electrical phenomena of the skin, *Psychol. Bull.*, 1929, 26, 64–115.

[2] Radecki, W., Recherches experimentales sur les phenomenes psychoelectriques, *Arch. de Psychol.*, 1911, 11, 209–293.

human subjects was that there was a constant and gradual drop in the resistance which seemed to be hastened by the breathing etc. But when a certain 'base level' (which varied in each subject) was reached the factors which had previously affected the resistance no longer had any influence. This may have been a part of the phenomenon of 'continued lowering of resistance' as described by Thouless.[3]

This continually decreasing resistance seemed to be markedly accelerated if a tourniquet were applied to the arm in such a way as to cause considerable distention of the blood vessels of the hand to which the electrodes were attached. For this purpose we used a sphygmometer which caused no disturbance from traction or muscular movement. With the blood vessels thus distended deep breathing produced very wide and sudden swings in the galvanometer needle. In this instance also we were unable to reproduce this phenomenon after the resistance had reached what we choose to call a 'base level.' It is possible that polarization played some part in this (we were using nickel electrodes at the time) for we were unable to elicit similar drops with a non-polarizing electrode.

It became evident that the best way to eliminate questionable factors and to get down to purely autonomic influence was to make use of experimental animals where the autonomic could be altered to suit our purposes.

Dr. Kuntz of the Neural Anatomy Department very kindly gave us the use of several cats whose inferior cervical ganglion had been removed on one side only. This ganglion had been so removed that no sympathetic fibers whatever remained connected with one of the front paws while the fibers to the other front paw were not disturbed. Such an arrangement gave us an excellent means for comparing the resistance under identical conditions in all save the autonomic supply. Even the question of polarization and other physical factors which might enter could not destroy or interfere with the results which would be taken from a comparative point of view.

The hairs on both front paws were completely removed

[3] Thouless, R. H., Causes of continued change in resistance, *Brit. J. Psychol.*, 1927, 17, 5–14.

with sodium sulphide and electrodes which had been made especially for this purpose were securely fastened to the flexor and extensor surfaces of each paw. These electrodes consisted of small copper plates to which copper wires had been soldered. They were made small enough to include only the 'cushion' part of the paw since this part alone contains sweat glands. They were kept moistened with gauze saturated with normal saline solution.

In the first series of experiments the cats were not given ether but after being placed comfortably in the arms of an assistant and gently stroked they became quickly resigned to the unusual apparatus and, although seldom absolutely free from some excitement, made few attempts to get rid of this incumbrance.

Readings were then obtained of the resistance of each paw. A Leeds Northrop resistance box with multiple resistance coils and a sensitive galvanometer were used to measure this resistance. A three-way switch was used at first but soon discarded on account of the fact that too many new points of resistance were thereby introduced. Direct connections were used in all the recorded experiments.

After taking readings every two or three minutes for 30 minutes or so it was found that the resistance fluctuated considerably on the sound side while the resistance of the operated paw remained absolutely the same. The cat was of course continually becoming restless and had to be quieted but whether quiet or restless the operated paw showed not the slightest change in resistance.

Marked emotional states were then produced by pinching and twisting the tail of the cat. In these states of extreme excitement the sound paw went through rapid and wide fluctuations in resistance but there was still no change whatever in the operated paw. The paws were firmly held so that the changes were not due to muscular movements. If this had been the case there is no reason why the operated paw should not have produced the same or at least some deflections of the needle. The same experiment was repeated on other cats similarly operated, that is with the inferior cervical gang-

lion completely removed on one side. In every case the sympathectomized side showed no change at all in resistance.

TABLE NO. 1

Unoperated Paw	Operated Paw (corresponding readings)
1,400 ohms (excited)	700 ohms
2,500 ohms, fairly quiet	700 ohms
900 ohms, slightly excited	700 ohms
1,300 ohms, fairly quiet	700 ohms
1,600 ohms, excited	700 ohms
2,100 ohms, markedly excited	700 ohms
(Period of rest)	
600 ohms, quiet	700 ohms
1,800 ohms, excited	700 ohms
1,000 ohms, somewhat quiet	700 ohms

Note—Only a few minutes elapsed between readings.

In the next series of experiments these cats were anæsthetized and after getting readings from both paws, 1 cc. of a 1–10,000 solution of Adrenalin Chloride was given subcutaneously to the first cat. Readings were taken every 4 or 5 minutes for 45 minutes. There was practically no change in the resistance of either paw. If the adrenalin subcutaneously were effective it was expected that changes in the resistance of the intact side would be produced while the operated side would remain the same, just as in the cases of emotional excitation. This was not the case however for there were slight variations in the resistance of both paws, the operated as well as the unoperated. The readings from the unoperated side ranged from 350 ohms to 440 ohms, while the range on the operated side was from 410 to 480 ohms.

It may be that the adrenalin when given subcutaneously had some effect on the muscles directly, but it seems more likely that these slight changes were more or less accidental, for on removing the electrodes after the experiment it was found that on both sides they had become somewhat loosened, consequently the contacts were not good. This had not happened in any of the other experiments. This explanation seems all the more likely in view of subsequent experiments where subcutaneous adrenalin produced no change on the operated side while the sound side showed considerable fluctuation.

The adrenalin was then given intravenously to another cat with very good results. One cc. of a 1–10,000 solution was again used. Here again the readings from the sympathectomized side showed no fluctuations whatever while the other side changed considerably. Within five minutes after the injection of the adrenalin intravenously the resistance on the sound side went rapidly from 620 ohms to a maximum of 980 ohms, after which it began to drop slowly until a half hour after the injection of the adrenalin the resistance on this side was 870 ohms. During all this time the readings from the sympathectomized paw remained absolutely the same, that is the resistance was 470 ohms before the ether was given, 470 ohms after ether but before adrenalin and 470 at every reading in the half hour after adrenalin.

TABLE NO. 2

Unoperated Paw	Operated Paw
400 ohms before ether.....	470 ohms before ether
620 ohms just before adrenalin.....	470 ohms after ether
980 ohms 5 min. after adrenalin.....	470 ohms 5 min. after adrenalin
960 ohms 20 min. after adrenalin.....	470 ohms 20 min. after adrenalin
870 ohms 25 min. after adrenalin.....	470 ohms 25 min. after adrenalin

It might be well to note that the blood vessels in the sympathectomized paw were, as one would expect, markedly dilated and remained in this condition. As a consequence the blood volume of the operated side remained much more constant than the side still controlled by the sympathetic. It has been pointed out by Uhlenbruck [4] that blood volume changes produce resistance changes. It does not seem however that constant blood volume alone accounts for the extraordinary immutability in resistance, for it is doubtful if the blood volume of the sympathectomized side remained so unvarying. Other factors which occurred during the experiment should have influenced the blood volume of both paws. For example, at times the paws were sharply elevated, at other times somewhat pendent; sometimes the assistant holding the cat was exerting considerable pressure on the proximal part of the leg, while at other times it was perfectly free. In

[4] Uhlenbruck, P., Uber den Autonomen Reflex, *Zeitschr. f. Biol.*, 1924, **81**, 51–56.

spite of all these changes the readings from the operated paw never changed so much as an ohm.

Gildmeister[5] thinks that the psychogalvanic reflex is merely a part of the general autonomic response. The results of our experiments seem to show that the autonomic nervous system plays a tremendously important part, if not the only part, in the psychogalvanic reflex. The remarkably constant readings from the sympathectomized paw could hardly be due to stability in the permeability of the cell membranes, which would otherwise be regulated by the autonomic as Gildmeister at one time postulated. The influence appears to be much more direct. The same may be said with regard to the sweat glands for although they are regulated by the sympathetic still in the case of cats with the inferior cervical ganglion removed, entire absence of variation in resistance seems difficult to explain by the absence of sweat gland activity alone. It seems difficult to explain for this reason, namely that the experiments were performed on cats at longer or shorter intervals after the sympathectomy. Use was made of one cat so soon after the sympathetic operation that he was still stuporous from the effects of amytol which had been given for that operation. At least in this last case it seems likely that there were some kind of changes going on in the sweat glands, sufficient to make a difference in resistance of a few ohms.

The findings here presented have not been checked from as many angles as the author would wish, nevertheless the influence of the autonomic on the psychogalvanic reflex may be seen from the above cited experiments even without more profound and exhaustive investigation.

Since the cats used were all from the experimental laboratory of Dr. Albert Kuntz I feel that we can be reasonably certain that the condition of the autonomic was in every instance as stated in the beginning of this paper.

[5] Gildmeister, M., Uber die elektrischen Leitungseigenschaften der Saugerhaut, *Pflüger's Arch. f. d. g. Physiol.*, 1922, **194**, 323–325.

(Manuscript received August 6, 1931)

THE REGULATED ALTERNATING CURRENT AS A TIME MEASURE

BY THOS. H. HOWELLS
University of Colorado

Within the last two or three years a new and improved means of time measurement has come into common usage by the general public. The synchronous electric clock is the agency which supplies this nearly-perfect record of the flight of time. The chief virtue, however, does not lie in the clock. The latter consists merely of an electric motor of the well-known synchronous type and a suitable system of reducing gears. The mechanism operates in unison with the generator at the power station just as if it were directly connected to it. The possibility of using such an electric motor as a clock lies in the fact that the generators in the power house are now operated in synchronism with a very precise chronometer, the accuracy of which is continually checked by astronomical observations. Thus it is that the average man can now have, for five dollars, a clock which is more accurate than the chronometer in the jewelry store. It seems obvious that such an accurate time indicator should find use in the psychological laboratory.

In addition to the element of accuracy there is another advantage of using the synchronous motor as a time reference in the laboratory, namely that it is much more rugged and powerful than the old-style clock motor and its accuracy is therefore not interfered with by any reasonable load that is put upon it. Any machine which is operated by a synchronous motor using regulated current moves with a perfectly constant speed and is therefore, in itself, an accurate time keeper.

The simplest method of making use of the regulated current in the laboratory is to purchase a simple electric clock and adapt if for the purposes at hand. There are several wheels or gears in the electric clock which run much faster than in the

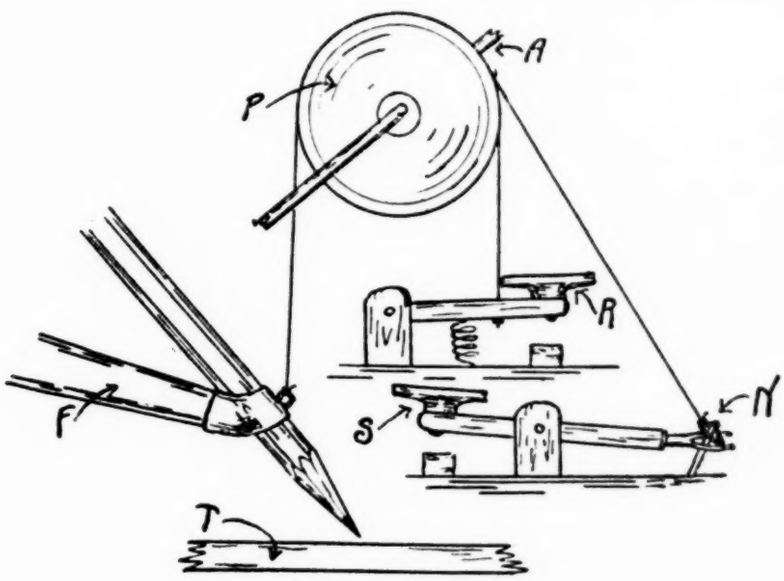
ordinary clock and these can be arranged to make electric contact so as to operate a time-marker or sounder at almost any frequency desired. No especial precaution need be taken to prevent friction at the points of contact as the motor does not slow down under a load as do most spring-operated devices. The motor in an electric clock is small but it is comparatively much more powerful than the spring motor of a clock. The writer has used a clock which has a starting wheel projecting at the back which rotates at a speed of one revolution per second. Attachment to this wheel of a metal arm which makes contact with a mercury cup provides an excellent substitute for the seconds-pendulum which is ordinarily used for operating an electric sounder. Lesser intervals can be obtained by attaching additional arms to this wheel, or by making connection with some of the gears that are turning more rapidly. Longer time intervals can be obtained by connecting with the seconds-hand, minute-hand or hour-hand on the face of the clock.

Larger types of synchronous motors can be used to operate special laboratory devices, such as a kymograph or tachistoscope. Once the proper speed-changing mechanism is provided for connecting the motor to the apparatus one can be quite certain that it will always move at a speed which is constant and accurate beyond any ordinary requirement.

One of the questions that naturally arises in regard to any new means of time measurement is whether it may make possible a simpler and more accurate type of chronoscope. The schema of such an instrument is shown in Fig. 1. T is the turn-table of a phonograph which is operated by a synchronous motor using the regulated current. The particular turntable motor used in the construction of this apparatus is part of a Rotrola phonograph pick-up unit. It is intended to be used in connection with a radio amplifier and speaker for the purpose of reproducing phonograph records. It is made by the Allen-Hough Company of Milwaukee. Any other similar unit would serve equally well provided it includes a synchronous motor. On the turn-table is placed a disc of paper which has been cut to the pattern of a phonograph record.

The inductor unit, which is ordinarily at the end of the flexible arm, *F*, is replaced by a pencil as shown in the diagram. This pencil is supported slightly above the paper by a cord which runs over the pulley, *P*, and is attached to a metal ring which is hooked over a metal peg at *N*. A branch from this cord is attached to the response key, *R*.

When the stimulus key, *S*, is depressed the cord is pushed off the peg which allows the pencil to drop on the rotating arm. It continues to draw a line on the paper until the pencil is lifted again by depression of the response key, *R*. The length of the line should be measured in terms of degrees of a circle



with a protractor. When the speed of rotation of the turntable is eighty revolutions per minute the degrees of rotation per second is 480, which means that each degree represents 2.08 sigma.

Additional records can be made on the same paper simply by shifting the pulley, *P*, laterally on the axis, *A*.

Special paper records are being constructed which are engraved with radial lines which give the sigma values directly. The paper blanks which are used on the disc type of recording thermometer are also very easily adapted for use with this apparatus.

It is possible, of course, to locate the keys in any position and at any distance from the instrument, provided that the

cord does not stretch a great deal under changes in tension. A means of electro-magnetically releasing and lifting the pencil-arm could be devised quite readily and might facilitate control and reduce mechanical errors to some extent.

Experience with this instrument seems to indicate that it is more accurate, and quicker and simpler in action than any of the less-expensive chronoscopes. The motor and pick-up units can be bought for about twenty-five dollars and may serve a variety of other purposes in the laboratory, such as those mentioned in the first part of this note.

(Manuscript received June 6, 1931)

A FLEXIBLE APPARATUS FOR RECORDING READING REACTIONS

BY MILES A. TINKER
University of Minnesota

This apparatus for exposing stimuli and recording responses is the same general scheme as described by Dodge [1] and Miles,[2] but with modifications in certain units which (1) permit wider application of the apparatus to problems involving visual perception; (2) increase the accuracy and ease of scoring records; and (3) reduce accompanying noises. A complete description of the Dodge apparatus, as modified by Miles, has been given by Tinker.[3]

In the present apparatus the exposure device was constructed almost entirely of aluminum and contains a window $1\frac{1}{4}$ by 4 inches. (A still larger window may be used.) This is large enough to expose paragraphs 30 words in length when printed in typewriter script or in 12 point type. Hence it is possible to expose such printed stimuli as single words, sentences, paragraphs, mental and scholastic test items, number series, formulae, geometric forms, and the like. The key for breaking the circuit to make the exposure may be manipulated by either the experimenter or the subject. Furthermore the response may be made by speaking into the voice key or by pressing a break key. In case the subject is employing his hands to facilitate computation (using pencil and paper) of the problem exposed in the window, the experimenter can break the circuit to record a response simultaneously with the writing of the response by the subject.

The timing unit is a Zimmerman vibrating metal bar which makes and breaks the time circuit by means of a mercury-dip

[1] R. Dodge and F. G. Benedict, Psychological effects of alcohol, *Carnegie Institute of Wash. Publ.*, 1915, No. **232**, 90–108.

[2] W. R. Miles, Alcohol and human efficiency, *Carnegie Institute of Wash. Publ.*, 1924, No. 333, 87–89.

[3] M. A. Tinker, Reading reactions for mathematical formulæ, *J. Exper. Psychol.*, 1926, **9**, 444–467.

contact. This timer may be adjusted for producing 5, 10, or 25 breaks per second. In reactions to words it is necessary to use a unit of one-fifth or one-tenth second to assure accurate and easy reading of the records, especially if the chronograph motor varies in speed. Such records may be read to one-twentieth second with a high degree of accuracy. For longer reaction times (reading paragraphs) where time intervals of $\frac{1}{2}$ or of one second are desired, a metronome or a time clock may be used in the time line circuit.

The recording unit of the apparatus is a Gaertner tape recording chronograph carrying two electro-magnetic markers. These markers (as used for time line) respond accurately with either 5 or 10 excursions per second. With two or more excursions per second, however, they become decidedly noisy which tends to become disturbing. To eliminate this noise a small frame was attached to the covering of the magnets. One end of this frame projected out, parallel to the marker arm. On this projection was a set screw with a soft rubber tip. With proper adjustment of the screw the rubber tip serves as a bumper which prevents the marker arm from clicking against the magnet. This reduces the excursion of the marker slightly, but eliminates the noise to such a degree that it is not noticed above the low hum of the chronograph motor. Any kymograph with a spiral drive may be substituted for the Gaertner chronograph.

(Manuscript received April 1, 1932)

PSYCHOLOGICAL REVIEW PUBLICATIONS

Original contributions and discussions intended for the Psychological Review should be addressed to

Professor Howard C. Warren, Editor PSYCHOLOGICAL REVIEW,
Princeton University, Princeton, N. J.

Original contributions and discussions intended for the Journal of Experimental Psychology should be addressed to

Professor Samuel W. Fernberger, Editor JOURNAL OF EXPERIMENTAL PSYCHOLOGY,
University of Pennsylvania, Philadelphia, Pa.

Contributions intended for the Psychological Monographs should be addressed to

Professor Herbert S. Langfeld, Editor PSYCHOLOGICAL MONOGRAPHS,
Princeton University, Princeton, N. J.

Reviews of books and articles intended for the Psychological Bulletin, announcements and notes of current interest, and *books offered for review* should be sent to

Professor Edward S. Robinson, Editor PSYCHOLOGICAL BULLETIN,
Institute of Human Relations, Yale University, New Haven, Conn.

Titles and reprints intended for the Psychological Index should be sent to

Professor Walter S. Hunter, Editor PSYCHOLOGICAL INDEX,
Clark University, Worcester, Mass.

All business communications should be addressed to

Psychological Review Company, Princeton, New Jersey

The

Journal of Experimental Psychology

is indexed in the

International Index to Periodicals

to be found in most public and college libraries

DIRECTORY
OF
AMERICAN PSYCHOLOGICAL PERIODICALS

American Journal of Psychology—Ithaca, N. Y.; Cornell University.
Subscription $6.50. 624 pages annually. Edited by M. F. Washburn, K. M. Dallenbach, Madison Bentley, and E. G. Boring.
Quarterly. General and experimental psychology. Founded 1887.

Journal of Genetic Psychology—Worcester, Mass.; Clark University Press.
Subscription $14.00 per yr.; $7.00 per vol. 1,000 pages ann. (2 vols.). Edited by Carl Murchison.
Quarterly. Child behavior, animal behavior, comparative psychology. Founded 1891.

Psychological Review—Princeton, N. J.; Psychological Review Company.
Subscription $5.50. 540 pages annually. Edited by Howard C. Warren.
Bi-monthly. General psychology. Founded 1894.

Psychological Monographs—Princeton, N. J.; Psychological Review Company.
Subscription $6.00 per vol. 500 pages. Edited by Herbert S. Langfeld.
Without fixed dates, each issue one or more researches. Founded 1895.

Psychological Index—Princeton, N. J.; Psychological Review Company.
Subscription $4.00. 300–400 pages. Edited by Walter S. Hunter and R. R. Willoughby.
An annual bibliography of psychological literature. Founded 1895.

Psychological Bulletin—Princeton, N. J.; Psychological Review Company.
Subscription $6.00. 720 pages annually. Edited by Edward S. Robinson.
Monthly (10 numbers). Psychological literature. Founded 1904.

Archives of Psychology—New York, N. Y.; Columbia University.
Subscription $6.00. 500 pages per volume. Edited by R. S. Woodworth.
Without fixed dates, each number a single experimental study. Founded 1906.

Journal of Abnormal and Social Psychology—Eno Hall, Princeton, N. J.; American Psychological Association.
Subscription $5.00. 448 pages annually. Edited by Henry T. Moore.
Quarterly. Abnormal and social. Founded 1906.

Psychological Clinic—Philadelphia, Pa.; Psychological Clinic Press.
Subscription $3.00. 288 pages. Edited by Lightner Witmer.
Without fixed dates (9 numbers). Orthogenics, psychology, hygiene. Founded 1907.

Psychoanalytic Review—Washington, D. C.; 3617 10th St., N. W.
Subscription $6.00. 500 pages annually. Edited by W. A. White and S. E. Jelliffe.
Quarterly. Psychoanalysis. Founded 1913.

Journal of Experimental Psychology—Princeton, N. J.; Psychological Review Company.
Subscription $7.00. 700 pages annually. Edited by Samuel W. Fernberger.
Bi-monthly. Experimental psychology. Founded 1916.

Journal of Applied Psychology—Baltimore, Md.; Williams & Wilkins Company.
Subscription $5.00. 400 pages annually. Edited by James P. Porter.
Bi-monthly. Founded 1917.

Journal of Comparative Psychology—Baltimore, Md.; Williams & Wilkins Company.
Subscription $5.00 per volume of 450 pages. Three volumes every two years. Ed. by Knight Dunlap and Robert M. Yerkes. Founded 1921.

Comparative Psychology Monographs—Baltimore, Md.; The Johns Hopkins Press.
Subscription $5.00. 400 pages per volume. Knight Dunlap, Managing Editor.
Published without fixed dates, each number a single research. Founded 1922.

Genetic Psychology Monographs—Worcester, Mass.; Clark University Press.
Subscription $14.00 per yr.; $7.00 per vol. 1,000 pages ann. (2 vols.). Edited by Carl Murchison. Monthly. Each number one complete research. Child behavior, animal behavior, and comparative psychology. Founded 1925.

Psychological Abstracts—Eno Hall, Princeton, N. J.; American Psychological Association.
Subscription $6.00. 700 pages ann. Edited by Walter S. Hunter and R. R. Willoughby.
Monthly. Abstracts of psychological literature. Founded 1927.

Journal of General Psychology—Worcester, Mass.; Clark University Press.
Subscription $14.00 per yr.; $7.00 per vol. 1,000 pages ann. (2 vols.). Edited by Carl Murchison.
Quarterly. Experimental, theoretical, clinical, historical psychology. Founded 1927.

Journal of Social Psychology—Worcester, Mass.; Clark University Press.
Subscription $7.00. 500 pages annually. Ed. by John Dewey and Carl Murchison.
Quarterly. Political, racial, and differential psychology. Founded 1929.